

**Estimation du risque résiduel d'infection d'un lot de plasma de
fractionnement VIH-infectieux à la suite de l'application de
technologies de réduction des pathogènes et l'abolition du critère
HARSAH.**

par

Eliana Aubé

mémoire présenté au Département de mathématiques en vue de l'obtention
du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, juillet 2021

À MES ENFANTS AÏSHA ET MAËL.

Le 15 juillet 2021

Le jury a accepté le mémoire de Madame Eliana Aubé dans sa version finale.

Membres du jury

Professeur Félix Camirand Lemyre

Directeur de recherche

Département de mathématiques

Professeur Antoine Lewin

Co-directeur de recherche

Faculté de Médecine et des Sciences de la Santé

Professeur Taoufik Bouezmarni

Membre interne

Département de mathématiques

Professeur Klaus Herrmann

Président-rapporteur

Département de mathématiques

SOMMAIRE

Au Canada, comme ailleurs, plusieurs critères d'admissibilité au don de sang ou de plasma ont été mis en place afin de limiter le risque de contamination du produit sanguin par des agents pathogènes et ainsi assurer la sécurité des receveurs. Un de ces critères vise les hommes ayant des relations sexuelles avec d'autres hommes (HARSAH). Ces derniers sont proscrits du don pour les trois mois suivant leur dernière relation sexuelle avec un autre homme. Cette exclusion du don est liée à l'incidence du VIH et à la prévalence du VIH dans la population HARSAH qui sont plus élevées que dans la population générale. Toutefois, le critère d'exclusion HARSAH est remis en question puisque les technologies de dépistage se sont extrêmement efficaces et qu'il est considéré comme discriminatoire. L'objectif de cette recherche est d'estimer, *in silico*, le risque résiduel qu'un lot de plasma, entendu ici comme la combinaison de plusieurs dons de plasma soumis au processus d'inactivation des pathogènes, soit contaminé par le VIH à la suite de l'abolition du critère d'exclusion visant les HARSAH.

Afin de bien évaluer le risque, un modèle basé sur la théorie des réseaux bayésiens a été construit et utilisé en combinaison avec une procédure de simulation de type Monte Carlo. Ce modèle permet d'évaluer le risque en tenant compte de la charge virale des dons infectés et de la technologie d'inactivation des pathogènes utilisée. Les dons formant les

lots de plasma sont générés individuellement de manière stochastique en se basant sur les caractéristiques du donneur. La distribution de ces dernières a été déterminée sur la base de données fournies par Héma-Québec et par la société canadienne du sang, d'un jeu de données américain de dons positif au VIH, ainsi que sur des éléments de connaissances tirés de la littérature. Le risque a été évalué suivant deux hypothèses principales, soit selon un délai d'exclusion HARSAH de trois mois (cas actuel) et selon une hypothèse permettant aux HARSAH sexuellement actifs de faire des dons de plasma. Plusieurs scénarios ont été simulés pour chacun des modèles pour tenir compte de l'incertitude des valeurs de certains paramètres et 300 000 lots de plasma ont été simulés pour chacun des scénarios.

Pour déterminer si un lot est considéré infectieux, un seuil typiquement considéré comme conservateur d'une copie d'ARN VIH par lot a été utilisé. Pour toutes les simulations, le risque d'obtenir un lot infectieux a été de 0 sur 300 000 lots (IC 95% [0 ; 0,0000123]) avec une valeur p plus petite que $2,2 \times 10^{-16}$. Selon le résultat des simulations, le risque résiduel lié à un changement à la politique d'exclusion du don visant les HARSAH est négligeable. On peut croire qu'il serait donc sécuritaire de permettre aux HARSAH sexuellement actifs de donner du plasma.

REMERCIEMENTS

À l'issue de ce travail de recherche et de rédaction, j'exprime toute ma reconnaissance à tous ceux qui, d'une façon ou d'une autre, ont contribué à la réalisation de ce mémoire.

Je souhaite d'abord exprimer ma gratitude envers mes directeurs de recherche Félix Camirand Lemyre et Antoine Lewin pour la confiance qu'ils m'ont accordée en acceptant de me confier ce projet de recherche. Merci de vos multiples conseils, de votre encadrement, de vos commentaires éclairés et de votre aide tout au long de ma maîtrise.

Je souhaite remercier Yves Grégoire pour son aide, pour ses réponses à mes multiples questions ainsi que pour le temps qu'il a pris pour réviser mes travaux. Je remercie aussi Olivier Germain qui a révisé et optimisé le code de mes simulations et sans qui les simulations auraient pris beaucoup plus de temps !

Je remercie le groupe "*Surveillance, Risk Assesment and Policy*" pour les nombreuses discussions visant l'élicitation des paramètres ainsi que l'organisme Héma-Québec et les gouvernements fédéral (Santé Canada) et provincial (Fond de recherche du Québec) pour le financement.

Toute ma considération va à Matthieu. Merci de m'avoir accompagnée dans mon cheminement. Ton support indéfectible et tes multiples encouragements ont, sans aucun doute, contribué à l'arrivée à terme de ce projet.

Finalement, je souhaite adresser un grand merci à mes parents, grands-parents et amis pour avoir cru en moi tout au long de mon parcours de retour aux études.

Eliana Aubé

Sherbrooke, juillet 2021

TABLE DES MATIÈRES

SOMMAIRE	iii
REMERCIEMENTS	v
TABLE DES MATIÈRES	vii
LISTE DES FIGURES	x
INTRODUCTION	1
CHAPITRE 1 — Différents modèles de quantification du risque résiduel	7
1.1 Revue de littérature	7
1.1.1 Modèle canadien 2003	7
1.1.2 Modèle américain	9
1.1.3 Modèle anglais	11
1.1.4 Modèle collaboratif 3-mois	12
1.2 Utilité et différence des modèles	14

1.3	Nécessité d'un nouveau modèle	16
CHAPITRE 2 — Outils mathématiques		18
2.1	Réseaux bayesiens	19
2.1.1	Notion de d-séparation	21
2.1.2	Distribution de probabilités et indépendance conditionnelle	24
2.2	Diagnostic de normalité	26
2.2.1	Diagramme quantile-quantile	26
2.2.2	Test de Shapiro-Wilk	28
2.3	Estimation non paramétrique de la densité	29
2.3.1	Choix du paramètre de lissage	30
2.3.2	Estimateur de densité à noyau log-normal	36
2.4	Évaluation du risque	38
2.4.1	Exemple	39
2.4.2	Probabilités conditionnelles	42
2.4.3	Test d'égalité des variances	44
CHAPITRE 3 — Application		46
3.1	Introduction	46
3.2	Le modèle	47
3.2.1	Profil du donneur	48

3.2.2	Probabilité d'infection	56
3.2.3	Charge virale	60
3.2.4	Test d'acide nucléique et technologie de réduction des pathogènes .	67
3.2.5	Différents scénarios	71
3.2.6	Résultats	75
3.3	Analyse de sensibilité	79
CONCLUSION		83
Annexe A		86
BIBLIOGRAPHIE		118

LISTE DES FIGURES

2.1	Réseau bayésien associé à l'exemple sur la grippe.	20
2.2	Densité de la charge virale.	31
2.3	GOA associé au risque d'infection d'un don.	42
3.1	GOA associé au profil du donneur.	53
3.2	GOA associé à la probabilité d'infection.	56
3.3	Histogramme de la charge virale.	62
3.4	Diagramme quantile-quantile de la charge virale.	63
3.5	Densité de la charge virale basée sur les données non catégorisées.	64
3.6	Densité de la charge virale par catégories.	66
3.7	GOA associé au TAN.	68

INTRODUCTION

Certaines conditions médicales nécessitent des transfusions sanguines ou encore des produits ou médicaments obtenus à partir de composants sanguins. Ces derniers sont obtenus par des collectes de sang ou de plasma. Ces collectes sont très importantes dans le domaine de la santé car elles permettent de répondre aux besoins en différents produits sanguins de certains patients. Au Québec, ces collectes sont effectuées par l'organisme Héma-Québec, alors que dans le reste du Canada, c'est la Société canadienne du sang qui s'emploie à cette tâche.

Afin d'assurer la qualité du produit ainsi que la sécurité des receveurs, l'admissibilité au don de sang et de plasma est conditionnelle au respect de certains critères. En effet, depuis l'identification du virus d'immunodéficience humaine (VIH), du virus d'hépatite B (VHB), du virus d'hépatite C (VHC) et d'autres infections transmissibles par transfusion, on a vu apparaître différents critères d'exclusion du don. Ces derniers visent à réduire le risque infectieux, qui est défini ici comme le risque qu'une infection se développe chez un individu suite à la contamination sanguine par un micro-organisme (virus, bactérie, parasite ou champignon) pathogène.

Parmi ces critères, on retrouve, entre autres, l'usage de drogues par voie intraveineuse

qui entraîne une exclusion permanente du don de sang. De plus, afin d'éviter de prélever des donneurs ayant été infecté (malaria, Chagas, Zika, etc.) suite à un voyage, une interdiction de 21 jours est actuellement applicable à tous donneurs revenant d'un pays autre que les États-Unis [39]. Ces derniers critères sont basés sur leur association à des comportements à risques. Cependant, un critère d'interdiction concernant les hommes qui ont des relations sexuelles avec d'autres hommes (HARSAH) est quant à lui basé sur l'orientation sexuelle des donneurs.

À travers la planète, les HARSAH ont longtemps été exclus du don de sang. Toutefois, depuis les années 90, par souci d'inclusion, les banques de sang tendent à vouloir retirer cette interdiction. Cependant, ceci doit-être effectué sans compromettre la sécurité transfusionnelle et sans augmenter le risque de transmission d'agents pathogènes. Il est donc nécessaire de bien évaluer les risques liés à ce changement. La mise en place de l'interdiction HARSAH au don de sang vient du fait que l'incidence, entendu ici comme le nombre de nouveaux cas dans la population sur un intervalle de temps choisi, et la prévalence, c'est-à-dire le nombre de personnes infectées dans la population à un moment déterminé, de certaines infections transmises sexuellement, dont le VIH, sont plus élevée au sein de ce groupe. En effet, la prévalence du VIH dans la population HARSAH est plus élevée que dans la population occidentale générale. Par exemple, à la fin de l'année 2016, l'Agence de la santé publique du Canada estimait que 49,1 % des personnes infectées par le VIH au pays étaient HARSAH [1].

La durée d'exclusion depuis la dernière relation sexuelle d'un homme avec un autre homme varie d'un pays à l'autre [40, 41]. Au Royaume-Uni, les HARSAH ont été exclus complètement du don de 1983 à 2011, puis la période d'exclusion est passée de 12 mois à trois mois en 2017. En Australie, les HARSAH font face à une exclusion de

trois mois depuis 2020. Cette exclusion était de 12 mois de 2005 à 2020. Les États-Unis avaient une interdiction de 12 mois depuis 2015, mais suite à l'avènement de la pandémie de coronavirus et dans un souci de maintenir un approvisionnement en produits sanguins suffisant, le critère d'exclusion HARSAH est passé à trois mois en avril 2020 [43]. La France est passée d'une exclusion permanente à 12 mois en 2016, puis a récemment modifié sa politique afin d'introduire une exclusion de quatre mois [42]. Certains pays comme la Grèce ont encore un critère d'exclusion permanente des HARSAH, alors que l'Italie et l'Espagne proposent plutôt une restriction liée au comportement à risque au lieu d'un critère visant les HARSAH spécifiquement.

Plus près de nous, en 2013, Santé Canada a autorisé l'abolition de l'exclusion permanente du don des HARSAH pour mettre en place une période d'exclusion temporaire : un délai d'exclusion du don de cinq ans depuis leur dernière relation sexuelle avec un homme. En 2016, ce délai d'interdiction est passé à 12 mois puis, plus récemment (juin 2019), à trois mois [38].

Lorsqu'un individu est infecté par un virus hématogène, des particules virales se retrouvent dans son organisme, en particulier dans son sang. Lorsque le taux de particules virales atteint un certain seuil, il est possible de détecter la présence du virus à l'aide de différents tests. La charge virale ou virémie est le terme qui désigne la quantité de particules virales présentes dans le sang de l'individu infecté. Le seuil de détection dépend du test utilisé. On appelle période silencieuse (ou période muette), l'intervalle de temps entre le moment de l'infection et la détection par le test de la présence du virus dans le sang. Il est important de considérer le risque lié à cette période, car il augmente la probabilité d'infection d'un échantillon. La charge virale durant cette période est inconnue et faible, mais bien présente [44].

Pour dépister la présence de virus tels que le VIH, il existe plusieurs types de tests, certains plus sensibles que d'autres. Il y a, entre autres, des tests de sérologie, qui permettent de détecter la présence d'anticorps au VIH dans le sang, et des tests d'amplification des acides nucléiques (TAN). Ce test détecte la présence d'acides nucléiques (ADN ou ARN) propres à un virus. Ils peuvent être effectués sur des dons individuels ou encore sur des lots constitués de plusieurs dons. Héma-Québec (HQ), la Société canadienne du sang (SCS) et plusieurs autres organismes de collecte de produits sanguins à travers le monde utilisent des TAN puisque ces derniers sont très sensibles. Ces tests permettent de réduire la période silencieuse puisqu'ils détectent la présence du virus avant l'apparition d'anticorps chez l'individu infecté [3]. Pour les TAN utilisés au Canada, la durée de la période silencieuse varie entre six et douze jours en fonction du test utilisé. Les tests peuvent être effectués individuellement sur chaque don, ou sur de petits bassins de dons.

La durée de la période silencieuse est directement liée à la sensibilité et à la spécificité du TAN utilisé. La spécificité est la probabilité d'obtenir un résultat négatif lors de l'absence du virus alors que la sensibilité est la probabilité d'obtenir un résultat positif lorsque l'échantillon est positif. Chaque test a une sensibilité et une spécificité qui lui est propre.

Il existe deux classes de produits sanguins obtenus par le don de sang ou de plasma. Il y a les produits labiles qui sont destinés à la transfusion au patient. Parmi ceux-ci, on retrouve les globules rouges, les plaquettes ainsi que le plasma thérapeutique. L'autre catégorie de produit concerne les produits stables. Il s'agit de médicaments dérivés du sang, tels que l'albumine, les facteurs antihémophiliques et différentes immunoglobulines. Le plasma destiné au fractionnement est principalement prélevé par aphérèse, mais peut

aussi provenir de dons de sang total. L'aphérèse est un procédé par lequel on prélève seulement le plasma. Ce dernier est séparé par centrifugation alors que le reste des composants sanguins sont réinjectés au donneur. C'est le fractionnateur qui transforme le plasma en produits stables. Pour ce faire, le fractionnateur reçoit les dons de plasma testés et les regroupe en lots. Dans certains cas, ces lots contiennent entre 4 000 litres et 6 000 litres de plasma [37].

L'objectif de la présente recherche est de quantifier le risque résiduel lié au don de plasma contaminé provenant de HARSAH dans le contexte de l'utilisation de la technologie de réduction des pathogènes (TRP) avant le fractionnement de plasma, suite à un changement au critère d'exclusion HARSAH. Le risque résiduel est défini comme le risque qui demeure par suite de l'application de mesures visant à réduire le risque inhérent à la situation. Dans le présent contexte, le risque résiduel est le risque de contamination qui subsiste à la suite de l'application du TAN et de la TRP dans le cas de dons de plasma faits par des HARSAH potentiellement porteurs du VIH.

Tel que discuté précédemment, plusieurs articles scientifiques présentent différents modèles de quantification de ce risque pour le don de sang sans l'application de la TRP. Au premier chapitre, quatre modèles seront présentés puisque ces derniers sont proches de la réalité canadienne et parce qu'ils évaluent des changements similaires.

Ensuite, au chapitre 2, les outils mathématiques utilisés dans le cadre de l'élaboration du modèle de simulation seront présentés.

Finalement, au troisième chapitre, les différentes étapes de simulation et ainsi que les

résultats ayant mené à l'évaluation du risque susmentionné seront discutés. Ce chapitre présentera aussi une analyse de sensibilité. Cette dernière consiste à faire varier certains des paramètres les plus susceptibles d'influencer le risque dans le but de vérifier la robustesse du modèle de simulation. Le chapitre présentera aussi une discussion sur les limites du modèle.

CHAPITRE 1

Différents modèles de quantification du risque résiduel

Afin de bien identifier les paramètres susceptibles d'avoir un impact sur l'évaluation du risque qu'un don de plasma soit VIH-infectieux et ainsi de bien construire le modèle, il est nécessaire de faire une revue de littérature sur les modèles antérieurs d'évaluation du risque associé au VIH lors d'un changement au critère HARSAH. Les quatre modèles se rapprochant le plus du contexte de cette recherche sont présentés à la prochaine section.

1.1 Revue de littérature

1.1.1 Modèle canadien 2003

Tout d'abord, au Canada en 2003, Germain et al.[5] ont présenté une méthode probabiliste. Leur modèle évalue le risque lié à l'éligibilité de don des hommes ayant des relations sexuelles avec d'autres hommes n'ayant pas eu de relations sexuelles avec d'autres

hommes dans les douze mois précédant le don. On calcule d'abord le nombre de nouveaux donneurs associé à ce changement de politique, noté ici N_{1an} . On a

$$N_{1an} = HARS AH_{tot} \times \mathbb{P}(\textit{éligible} \mid HARS AH) \times \mathbb{P}(\textit{don} \mid HARS AH, \textit{éligible}),$$

où $HARS AH_{tot}$ est la population totale de HARS AH, $\mathbb{P}(\textit{éligible} \mid HARS AH)$ la probabilité qu'un HARS AH devienne éligible au don et $\mathbb{P}(\textit{don} \mid HARS AH, \textit{éligible})$ la probabilité qu'un HARS AH nouvellement éligible fasse un don.

L'intuition de la formule ci-dessus est que $\mathbb{P}(\textit{éligible ET don} \mid HARS AH) = \mathbb{P}(\textit{éligible} \mid HARS AH) \mathbb{P}(\textit{don} \mid HARS AH, \textit{éligible})$. C'est donc seulement une application de propriétés de calculs de probabilités.

On calcule ensuite le nombre d'unités additionnelles contaminées qui se retrouveraient disponibles pour transfusion, noté U_{1an} :

$$U_{1an} = N_{1an} \times \mathbb{P}(VIH \mid \textit{nouveau donneur}) \times [\mathbb{P}(\textit{faux négatif}) + \mathbb{P}(\textit{non détecté} \mid \textit{variant}) + \mathbb{P}(PS) + \mathbb{P}(\textit{erreur tech}) + \mathbb{P}(\textit{erreur inventaire}) + \mathbb{P}(\textit{urgent})],$$

où $\mathbb{P}(VIH \mid \textit{nouveau donneur})$ est la probabilité d'infection des nouveaux donneurs, $\mathbb{P}(\textit{faux négatif})$ est la probabilité d'obtenir un faux négatif, $\mathbb{P}(\textit{non détecté} \mid \textit{variant})$ est la probabilité de non-détection d'échantillons dû à des souches virales variantes, $\mathbb{P}(PS)$ est la probabilité qu'un don soit fait durant la période silencieuse, $\mathbb{P}(\textit{erreur tech})$ est la probabilité d'avoir une erreur de laboratoire, $\mathbb{P}(\textit{erreur inventaire})$ est la probabilité qu'un échantillon VIH positif soit mal identifié et $\mathbb{P}(\textit{urgent})$ est la probabilité qu'une

transfusion soit effectuée avant les tests. Les sources d'erreurs sont indépendantes entre elles. Par exemple, la probabilité d'avoir une erreur de laboratoire n'a aucun effet sur la probabilité qu'un échantillon VIH positif soit mal identifié.

Dans ce modèle, la probabilité d'infection au VIH dépend du statut du donneur et si on considère $\mathbb{P}(\textit{erreurs})$ la somme des erreurs de la formule précédente, on a

$$\mathbb{P}(\textit{don infecté}) = \mathbb{P}(\textit{VIH} \mid \textit{nouveau donneur}) \times \mathbb{P}(\textit{erreurs}).$$

Alors, la probabilité qu'une unité provenant d'un HARSAH éligible soit contaminée par le VIH et se retrouve disponible pour la transfusion est : U_{1an}/N_{1an} .

Pour la simulation Monte Carlo de 10 000 itérations, une distribution *a priori* a été assignée à chacun des paramètres ($HARSAH_{tot}$, $\mathbb{P}(\textit{éligible} \mid HARSAH)$, $\mathbb{P}(\textit{don} \mid HARSAH, \textit{éligible})$, $\mathbb{P}(\textit{erreurs})$) en fonction des valeurs possibles. Des distributions triangulaires, binomiales et lognormales ont été utilisées en fonction des données initiales.

1.1.2 Modèle américain

Aux États-Unis, en 2009, Anderson et al.[6] ont aussi développé un modèle de simulation probabiliste afin de quantifier le risque lié à une diminution du temps d'exclusion du don des HARSAH à cinq ans et un an. Tout comme dans le modèle précédent, on calcule le nombre de donneurs additionnels N_{DM5} sous la condition choisie. Ici, seulement le cas du délai de cinq ans sera discuté. On a donc,

$$\begin{aligned} N_{DM5} = & N_{HA} \times \mathbb{P}(HARSAH \mid \textit{homme}) \times \mathbb{P}(5 \textit{ ans} \mid HARSAH) \\ & \times \mathbb{P}(\textit{donneurs}) - N_{HA} \times \mathbb{P}(HARSAH \mid \textit{homme}) \times \mathbb{P}(5 \textit{ ans} \mid HARSAH) \\ & \times \mathbb{P}(\textit{donneurs}) \times \mathbb{P}(\textit{don} \mid \textit{non-éligible}), \end{aligned}$$

où N_{HA} est le nombre d'hommes adultes (18-65) aux États-Unis, $\mathbb{P}(HARSAH \mid homme)$ est la probabilité d'être HARSAH si on est un homme, $\mathbb{P}(5 \text{ ans} \mid HARSAH)$ la probabilité de ne pas avoir eu de relations sexuelles avec d'autres hommes dans les cinq dernières années étant donné que l'homme est HARSAH, $\mathbb{P}(donneurs)$ est la probabilité d'être un donneur dans la population générale et $\mathbb{P}(don \mid non\text{-}éligible)$ est la probabilité qu'un HARSAH ayant eu au moins une relation sexuelle avec un autre homme au cours des cinq dernières années donne potentiellement du sang malgré l'interdiction.

En d'autres termes, on soustrait le nombre de donneurs non-compliants, c'est-à-dire les HARSAH qui donnent du sang sans se conformer à l'interdiction, au nombre de nouveaux donneurs potentiels.

Ensuite on estime le nombre d'unités de sang contaminées qui deviendraient disponible pour transfusion (N_{BSM5}) :

$$\begin{aligned} N_{BSM5} = N_{DM5} \times [& (\mathbb{P}(erreur \text{ TAN} \mid faux \text{ négatifs}) \\ & \times \mathbb{P}(erreur \text{ TDI} \mid faux \text{ négatifs})) + (\mathbb{P}(erreur \text{ quarantaine} \mid hôpitaux) \\ & \times \mathbb{P}(dons \mid hôpitaux) + \mathbb{P}(erreur \text{ quarantaine} \mid centre) \\ & \times (1 - \mathbb{P}(dons \mid hôpitaux)))], \end{aligned}$$

où $\mathbb{P}(erreur \text{ TAN} \mid faux \text{ négatifs})$ est la probabilité d'erreur du TAN dû à des faux négatifs, $\mathbb{P}(erreur \text{ TDI} \mid faux \text{ négatifs})$ est la probabilité d'erreur du test de dosage immuno-enzymatique dû à des faux négatifs, $\mathbb{P}(erreur \text{ quarantaine} \mid hôpitaux)$ est la probabilité d'erreur de mise en quarantaine dans les hôpitaux, $\mathbb{P}(dons \mid hôpitaux)$ est la proportion de dons faits dans les hôpitaux et $\mathbb{P}(erreur \text{ quarantaine} \mid centre)$ est la probabilité d'erreur de mise en quarantaine dans les centres de collecte de sang.

Différentes distributions ont été utilisées pour les différents paramètres du modèle : normale, bêta, triangulaire et uniforme. Puis, comme dans le modèle canadien, on a effectué une simulation Monte Carlo de 10 000 itérations.

1.1.3 Modèle anglais

En 2011, en Angleterre, Davison et al.[7] ont présenté un modèle déterministe. Ce dernier évalue le risque lié à différents scénarios : exclusion des HARSAH du don de sang à vie, exclusion de cinq ans et aucune exclusion. On calcule d'abord le risque lié à la période silencieuse (R_{PS}) par 100 000 dons :

$$R_{PF} = I_{an} \times L_{PS} \times S,$$

où I_{an} est l'incidence par 100 000 personnes-années, L_{PS} est la longueur de la période silencieuse (PS) en années et S est la longueur moyenne de l'intervalle entre les dons pour tous les donneurs divisée par la médiane de l'intervalle entre les dons pour les donneurs ayant subi la séroconversion entre les dons.

On calcule l'incidence du VIH comme suit : Tout d'abord, on calcule l'incidence chez les donneurs récurrents (I_{DR}) :

$$I_{DR} = \frac{\text{Nombre de séroconversions}}{\text{Nombre de dons répétés} \times \text{intervalle interdon moyen (années)}}.$$

Puis on calcule l'incidence totale par 100 000 personnes (I_{an}) :

$$I_{an} = (I_{DR} + I_{DR} \times \text{ajustement pour nouveau donneur}) \times 100\,000.$$

L'incidence dépend donc du statut du donneur (nouveau ou connu).

On calcule maintenant le risque lié à la sensibilité du test, il s'agit en fait du risque à un faux résultat négatif par 100 000 dons (R_{FN}). On a

$$R_{FN} = \frac{\text{Prévalence}}{\text{sensibilité du test}} \times \frac{1 - \text{sensibilité du test}}{1\,000}.$$

La prévalence du VIH se calcule comme suit :

$$\text{Prévalence} = \frac{\text{nombre de dons VIH positifs} + \text{Nombres de dons positifs supplémentaires attendus}}{\text{nombre de dons testés} + \text{nombre de dons supplémentaires testés}} \times 100\,000.$$

Ensuite, on calcule le risque associé au fait qu'un don infecté se retrouve disponible pour la transfusion (erreur de classement) dû aux erreurs de test et de traitement (R_{err}) :

$$R_{err} = \frac{\text{Prévalence} \times \text{fréquence des erreurs de processus}}{1\,000}.$$

Finalement, on obtient le risque que des unités contaminées par le VIH deviennent disponibles pour transfusion (R_{tot}) en effectuant la somme des risque calculés précédemment :

$$R_{tot} = R_{PS} + R_{FN} + R_{err}.$$

1.1.4 Modèle collaboratif 3-mois

En 2020, O'brien et al.[12] ont développé un modèle de simulation Monte Carlo avec paramètres déterminés afin d'estimer le risque résiduel de transmission du VIH dans

un scénario présentant une exclusion de trois mois suivant la dernière relation sexuelle HARSAH. On y calcule le risque résiduel comme suit :

$$R_{\text{résiduel}} = R_{\text{erreur test}} + R_{\text{sensibilité}} + R_{PS},$$

où R_{PS} est le risque associé à la période silencieuse, $R_{\text{sensibilité}}$ est le risque lié à la sensibilité du test TAN et $R_{\text{erreur test}}$ est le risque associé à l'erreur liée au test.

Afin de calculer le risque lié à la sensibilité du TAN

$$R_{\text{sensibilité}} = \text{Prévalence}_{VIH} \times \text{sensibilité},$$

et le risque lié aux erreurs de test,

$$R_{\text{erreur test}} = \text{Prévalence}_{VIH} \times \text{erreur de test},$$

on a besoin de calculer le taux de prévalence du VIH dans la population. Ce calcul tient compte du nombre de nouveaux donneurs séropositifs attendus ainsi que du nombre de nouveaux donneurs HARSAH attendus. La formule du calcul de la prévalence est donc :

$$\text{Prévalence}_{VIH} = \frac{N_{\text{positif}} + N_{\text{HARSAH}+}}{N_{\text{testés}} + N_{\text{HARSAHnouveaux}}},$$

où N_{positif} est le nombre de cas VIH positifs observés, $N_{\text{HARSAH}+}$ est le nombre de nouveaux cas HARSAH VIH positifs. Ce dernier varie selon une distribution de poisson centrée sur l'estimation initiale du nombre de nouveaux cas. 100 000 itérations sont effectuées afin d'obtenir 100 000 valeurs de prévalence. $N_{\text{testés}}$ est le nombre de dons testés

et $N_{\text{HARSAH}_{\text{nouveaux}}}$ est le nombre de dons faits par de nouveaux donneurs HARSAH (prévision).

Le risque lié à la période silencieuse est évalué avec la formule suivante :

$$R_{PS} = \frac{\frac{N_{\text{séroconversions}}}{N_{\text{dons}}} \times \mathbb{P}(\text{connus}) + I_{\text{connus}} \times \mathbb{P}(\text{nouveaux}) \times FC_{\text{nouveaux}}}{II_{\text{moy}}},$$

où $N_{\text{séroconversions}}$ est le nombre de séroconversions basé sur l'incidence du VIH dans la population de donneurs (HARSAH et non-HARSAH). N_{dons} est le nombre de dons, $\mathbb{P}(\text{connus})$ est basée sur la proportion de donneurs connus, I_{connus} est l'incidence du VIH chez les donneurs connus ($\frac{N_{\text{séroconversions}}}{N_{\text{dons}}}$), $\mathbb{P}(\text{nouveaux})$ est basée sur la proportion de nouveaux donneurs, $FC_{\text{nouveaux}} =$ est le facteur de correction de l'incidence pour les nouveaux donneurs et II_{moy} est l'intervalle interdon moyen.

Dans ce modèle, on suppose que l'incidence et la prévalence du VIH varient en fonction du statut du donneur (nouveau ou connu).

1.2 Utilité et différence des modèles

La modélisation mathématique est nécessaire dans l'évaluation du risque puisqu'elle permet d'orienter les réflexions, dans une certaine mesure, entourant l'occurrence d'un événement et donc de quantifier le risque avant l'instauration de mesures. Les modèles présentés ci-haut peuvent être classés par catégorie : les modèles déterministes, les modèles stochastiques ou encore une combinaison des deux.

Dans les modèles fréquentistes, on détermine une valeur fixe pour chacun des paramètres alors que les modèles bayésiens (modélisation stochastique) permettent d'inclure de la variabilité dans les paramètres. Les données nécessaires à l'évaluation du risque de transmission du VIH par don de sang ou encore par plasma de fractionnement sont très difficiles à obtenir. Habituellement, lorsque l'on cherche à évaluer un risque inhérent à une situation particulière, on effectue une collecte de données. Par exemple, dans le domaine de la santé, on procède souvent à des études cliniques afin d'obtenir un échantillon représentatif de la population étudiée. Dans le contexte de l'évaluation du risque de contamination lié aux dons de sang ou de plasma, il est pratiquement impossible d'effectuer un essai clinique, en raison notamment du fait que les produits sanguins se font rares et que l'approvisionnement est une préoccupation constante des banques de sang. De plus, il pourrait être difficile d'obtenir un échantillon représentatif de la population de donneurs. En effet, les méthodes de recrutement habituelles rejoignent facilement la population générale, mais plus difficilement certains sous-groupes comme les HARSAH. Ces derniers pourraient donc être faussement représentés ce qui mènerait à des résultats biaisés. Vu l'impossibilité de baser la présente évaluation du risque sur des essais cliniques, il convient de développer d'autres méthodes pour estimer ce dernier. Puisque plusieurs des valeurs utilisées dans la présente simulation sont obtenues par estimation, il devient pertinent d'opter pour un modèle stochastique.

Malgré l'impossibilité de récolter des données complètes menant à l'estimation directe du risque, il est possible de procéder par estimation à l'aide d'un modèle probabiliste puisque les connaissances des experts sur chacun des éléments ayant un impact sur le risque de contamination sont assez profondes.

1.3 Nécessité d'un nouveau modèle

Dans l'intention d'abolir la période d'exclusion du don de sang pour les HARSAH, il est nécessaire de bien évaluer et quantifier le risque associé à un tel changement. La Société canadienne du sang (SCS) et d'Héma-Québec veulent d'abord évaluer l'effet d'une modification au critère d'exclusion lié au plasma prélevé par aphérèse et destiné au fractionnement afin de appuyer la demande de changement. Ce processus implique l'application d'une technologie de réduction des pathogènes (PRT), technologie qui n'a pas été considérée dans les études antérieures de quantification du risque résiduel suite à une contamination d'un don au VIH. Afin de bien modéliser l'efficacité de cette technologie, il devient nécessaire de bien identifier les dépendances entre les paramètres afin d'associer une valeur de charge virale (VIH) aux dons infectés simulés. Cette dernière est estimée à partir de données recueillies auprès de donneurs infectés au VIH. L'objectif de la présente recherche est donc d'estimer, *in silico*, le risque d'obtenir un produit, issu du plasma de fractionnement, potentiellement contaminé au VIH suite à un changement au critère d'exclusion HARSAH pour le plasma récolté par aphérèse. Puisque le critère canadien d'exclusion HARSAH de trois mois est relativement récent et qu'il n'y a pas suffisamment de données récoltées depuis, il est nécessaire de simuler à partir de données recueillies ou estimées lorsque le critère d'exclusion HARSAH de 12 mois était en place. L'utilisation d'une approche basée sur des systèmes de structuration des connaissances comme les systèmes experts (SE) devient pertinente. Dans le cadre de la présente recherche, un tel SE aurait pour but de reproduire le raisonnement d'un ou plusieurs experts internationaux, principalement membres du groupe SRAP (Surveillance, Risk Assessment and Policy), en raisonnant sur des faits et des règles. Toutefois, ces derniers peuvent être incertains puisque basés sur des phénomènes naturels. Les réseaux bayésiens, qui s'apparentent en quelque sorte à des systèmes experts probabilistes, permettent de modéliser, de manière probabiliste, des systèmes complexes en se basant sur des connaissances liées aux in-

teractions locales entre les différents groupes de variables en jeu. C'est donc la raison pour laquelle, en utilisant des données fournies par la SCS et Héma-Québec et d'autres données fournies par des experts, il devient possible d'utiliser une approche Monte Carlo combinée à une modélisation par réseau bayésien de la distribution jointe des variables du système pour estimer le risque de contamination au VIH d'un lot de plasma traité à la TRP et destiné au fractionnement.

CHAPITRE 2

Outils mathématiques

Afin de bien simuler les phénomènes biologiques ou comportementaux, plusieurs outils mathématiques sont nécessaires. Ce chapitre présentera les outils utilisés dans la réalisation de la présente recherche.

Tel que mentionné ci-dessus, le modèle développé pour l'estimation du risque résiduel repose sur l'utilisation d'un réseau bayésien pour éliciter la distribution de probabilité jointe du système. Ce chapitre débutera donc par une introduction sommaire à ces objets probabilistes. Ensuite, des tests diagnostiques de non-normalité seront discutés, puisque ceux-ci ont été utilisés pour la modélisation de la distribution de la charge virale. Par la suite, l'estimation de densité par noyaux sera abordée, puisque le recours à cette technique a été nécessaire pour caractériser adéquatement la distribution de la charge virale. Enfin, le chapitre se termine sur une exposition de l'approche utilisée pour l'évaluation du risque par recours à un réseau bayésien.

2.1 Réseaux bayesiens

Établir une distribution de probabilité jointe sur des variables donnant lieu à un nombre élevé de résultats possibles peut paraître comme une tâche colossale. Considérons l'exemple présenté dans le livre *Probabilistic Graphical Models : Principles and Techniques* [36] :

Soit deux maladies, la grippe et le rhume des foins, les saisons de l'année et deux symptômes soit la congestion et la douleur musculaire. Les maladies peuvent se retrouver en même temps chez un patient et doivent donc être considérées comme deux variables distinctes. Il y a donc 8 variables qui peuvent mener à 64 résultats distincts.

Il pourrait être ardu de définir la distribution de probabilité de cet exemple. Les modèles graphiques probabilistes (MGPs) offrent la possibilité de décrire, de manière compacte, des distributions de probabilités complexes.

Précisément, les MGPs offrent une avenue de modélisation de distributions de probabilités conjointes qui allie la théorie des graphes à la théorie des probabilités. Intuitivement, un MGP associé à un système formé d'un ensemble de variables aléatoires \mathcal{X} est la donnée d'une composante graphique \mathcal{G} et d'une distribution de probabilité P telle qu'il soit possible de lire les relations de dépendance ou d'association que sous-tendent P au travers de la lecture du graphe. La composante graphique peut être un graphe orienté, non orienté, ou semi-orienté, au sein duquel les noeuds représentent les variables aléatoires du système, et les arêtes reflètent les relations de dépendance ou d'association directes, c-à-d non médiées par aucune autre variable du système. Un graphe est typiquement indexé par une paire $(\mathcal{X}, \mathcal{E})$, où \mathcal{X} est l'ensemble des noeuds du graphe (qui coïncide ici avec les variables aléatoires du système) et \mathcal{E} est l'ensemble des arêtes reliant des noeuds dans \mathcal{G} . Dans ce contexte, pour $X_1, X_2 \in \mathcal{X}$, il est commode d'écrire $X_1 \rightarrow X_2 \in \mathcal{E}$ s'il existe

une arête orientée qui part de X_1 pour aller vers X_2 dans \mathcal{G} , et $X_1 - X_2$ s'il existe une arête non orientée entre X_1 et X_2 dans \mathcal{G} . Typiquement, dans la littérature, on utilise des arêtes orientées pour refléter une relation de type causale entre deux variables, et des arêtes non orientées pour représenter des interrelations dites réflexives (chacune des deux variables est susceptible d'influencer l'autre).

Les réseaux bayésiens sont un type particulier de modèles graphiques probabilistes qui conviennent à la modélisation de systèmes au sein desquels toutes les relations entre les variables sont interprétées comme causales. Voici comme exemple le graphe du réseau bayésien associé à l'exemple de la grippe et du rhume des foins.

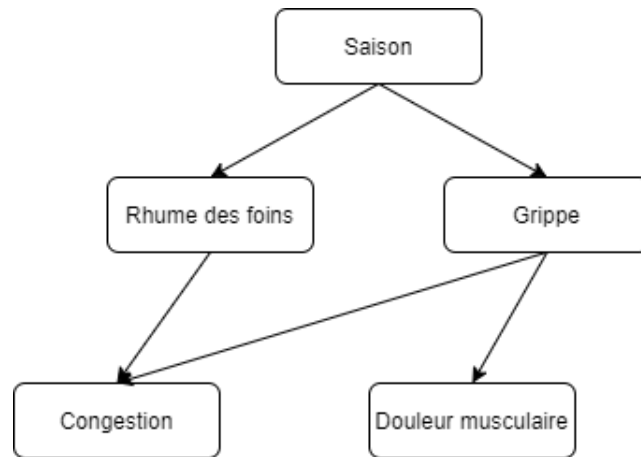


FIGURE 2.1 – Réseau bayésien associé à l'exemple sur la grippe.

La composante graphique est constituée d'un *graphe orienté acyclique* (GOA), c'est-à-dire un graphe où toutes les arêtes sont orientées et où on ne retrouve pas de chemin formant un cycle. Un chemin \mathcal{C} entre deux sommets $X_i, X_j \in \mathcal{X}$ d'un graphe orienté $\mathcal{G} = (\mathcal{X}, \mathcal{E})$ est une succession d'arêtes dirigées et de sommets entre les deux sommets initiaux. \mathcal{C} forme un cycle si ses extrémités coïncident et ses autres sommets ne se répètent pas. À noter qu'afin de permettre la représentation d'un schéma causal à l'aide d'un graphe, il est

nécessaire que ce dernier ne contienne pas de cycle. En effet, si c'était le cas, on aurait alors un patron du type $X_{i_1} \rightarrow X_{i_2} \rightarrow \dots \rightarrow X_{i_p} \rightarrow X_{i_1}$, et on interpréterait alors que X_{i_1} cause X_{i_2} , qui cause à son tour X_{i_3} , et ainsi de suite jusqu'à X_{i_p} qui a un effet causal sur X_{i_1} , ce qui contredit le principe de causalité.

Les réseaux bayésiens sont très appropriés pour l'évaluation du risque, car ils permettent de représenter certains systèmes complexes de manière modulaire, c'est-à-dire en spécifiant localement la structure de dépendance de petits groupes de variables du système, en se basant sur l'interprétation que se font des experts des interrelations présentes. Dans le contexte de cette recherche, la construction du réseau bayésien est basée sur l'élicitation, faite par des experts, d'une structure de dépendance entre les variables nécessaires à l'estimation du risque de contamination du plasma sanguin.

Plusieurs concepts associés à la théorie des réseaux bayésiens sont nécessaires pour bien saisir la manière de spécifier un tel MGP. On présentera quelques notions essentielles à la bonne compréhension de la définition d'un réseau bayésien telles que la notion de d-séparation, de distribution de probabilité et d'indépendance conditionnelle.

2.1.1 Notion de d-séparation

La définition d'un réseau bayésien repose sur la construction d'un cadre qui rend cohérent la manière de lire les relations de dépendances, ou plus précisément les indépendances, dans un graphe orienté avec celui fourni par la théorie des probabilités. En présence d'un graphe orienté acyclique, c'est le critère de d-séparation qui permet d'atteindre cette cohérence et ainsi définir la façon d'interpréter les indépendances (conditionnelles) dans le graphe [23, 24].

La connaissance de certains concepts est nécessaire afin de bien comprendre le critère de d-séparation. Définissons d'abord quelques termes utilisés dans la théorie des graphes.

Définition 2.1.1 (Graphe orienté acyclique (GOA)). *Soit $\mathcal{G} = (\mathcal{X}, \mathcal{E})$ un graphe orienté. On dit de \mathcal{G} qu'il est un **graphe orienté acyclique** s'il ne possède pas de cycles.*

Définition 2.1.2 (Descendant). *Soit $\mathcal{G} = (\mathcal{X}, \mathcal{E})$ un GOA. Un sommet $X_2 \in \mathcal{X}$ est un **descendant** d'un autre sommet $X_1 \in \mathcal{X}$ si et seulement s'il existe un chemin partant de X_1 vers X_2 .*

Introduisons maintenant la notion de chemin bloqué (ou inactif) qui est à la base de la notion de d-séparation.

Définition 2.1.3 (Chemin actif). *Étant donné un GOA $\mathcal{G} = (\mathcal{X}, \mathcal{E})$ et $X_1, \dots, X_k \in \mathcal{X}$, un chemin qui relie X_1 à X_k est dit **actif** étant donné un groupe de sommets $Z \subset \mathcal{X}$ si une des conditions suivantes est respectée pour $i \in 1, \dots, k-2$:*

1. *(Le chemin n'est pas linéaire)*

$$X_i \rightarrow X_{i+1} \rightarrow X_{i+2} \quad \text{ou} \quad X_i \leftarrow X_{i+1} \leftarrow X_{i+2} \quad \text{alors} \quad X_{i+1} \notin Z$$

2. *(Le chemin est divergent)*

$$X_i \leftarrow X_{i+1} \rightarrow X_{i+2} \quad \text{où} \quad X_{i+1} \in Z \quad \text{ou} \quad \text{un des descendants de } X_{i+1} \text{ est dans } Z$$

3. *(Le chemin est convergent)*

$$X_i \rightarrow X_{i+1} \leftarrow X_{i+2} \quad \text{où} \quad X_{i+1} \notin Z \quad \text{et} \quad \text{aucun des descendants de } X_{i+1} \text{ n'est dans } Z$$

*Un chemin est dit **inactif** s'il n'est pas **actif**.*

Considérons donc les trois variables X_1, X_2 et X_3 en tant que sommets d'un GOA. Il y a quatre chemins possibles pour aller de X_1 à X_2 en passant par une troisième variable X_3 :

1. $X_1 \longrightarrow X_3 \longrightarrow X_2$ (X_1 est une cause à X_2)
2. $X_1 \longleftarrow X_3 \longrightarrow X_2$ (X_3 est une cause à X_1 et X_2)
3. $X_1 \longleftarrow X_3 \longleftarrow X_2$ (X_2 est une cause à X_1)
4. $X_1 \longrightarrow X_3 \longleftarrow X_2$ (X_1 et X_2 sont une cause à X_3)

Examinons le cas où X_3 n'est pas observée. Les trois premiers chemins sont actifs puisque dans tous les cas, on peut identifier une dépendance entre X_1 et X_2 . Dans le dernier cas, il n'est pas possible de lier X_1 et X_2 , le chemin est donc inactif. Toutefois, lorsque X_3 est observée, on dit que X_3 bloque le chemin, qui devient donc inactif dans les trois premiers scénarios. On en arrive donc à la définition de d-séparation qui permet de caractériser les indépendances présentes dans un GOA.

Définition 2.1.4 (d-séparation). *Soit $\mathcal{G} = (\mathcal{X}, \mathcal{E})$ un GOA. Soit $X \subset \mathcal{X}$, $Y \subset \mathcal{X}$ et $Z \subset \mathcal{X}$ des ensembles de variables aléatoires disjoints dans \mathcal{X} . Alors, on dit que $X_i \in X$ et $Y_j \in Y$ sont **d-séparées** par Z si tous les chemins entre X_i et Y_j sont bloqués par Z et on note*

$$(X \perp\!\!\!\perp Y \mid Z)_d$$

Cette notion est utilisée pour construire le réseau bayésien et pour représenter graphiquement les dépendances entre certains paramètres. Elle est basée sur l'avis et les connaissances des experts. Ces derniers identifient quels paramètres en influencent d'autres, ce qui permet de caractériser les conditions qui permettent de traiter une information localement sans affecter tout le réseau.

Précisément, la notion de d-séparation mène à la caractérisation suivante des structures graphiques qui serviront à la construction de réseaux bayésiens : dans un GOA, un noeud X est d-séparé de ses non-descendants étant donné ses parents. Cette caractérisation sera formalisée dans ce qui suit.

2.1.2 Distribution de probabilités et indépendance conditionnelle

Si on utilise un graphe orienté acyclique afin de modéliser les relations causales directes entre des variables aléatoires, il est possible de décrire son schéma de décomposition afin de simplifier le calcul de probabilité. On considère donc un espace fini de variables aléatoires $\mathcal{X} = \{X_1, \dots, X_n\}$. On désire caractériser le comportement joint entre ces variables. La probabilité jointe de \mathcal{X} , $\mathbb{P}(\mathcal{X})$ comporte 2^{n-1} paramètres. Toutefois, si on connaît les dépendances entre les variables, on peut toujours écrire la distribution de probabilité comme suit :

$$\mathbb{P}(X_1, \dots, X_n) = \mathbb{P}(X_1) \mathbb{P}(X_2 | X_1) \mathbb{P}(X_3 | X_1, X_2) \dots \mathbb{P}(X_n | X_1, \dots, X_{n-1}),$$

ou, de manière succincte,

$$\mathbb{P}(X_1, \dots, X_n) = \mathbb{P}(X_1) \prod_{i=1}^n \mathbb{P}(X_i | X_{i-1}, \dots, X_1). \quad (2.1)$$

Intuitivement, le GOA associé à un réseau bayésien offre une manière de lire la présence d'indépendances conditionnelles entre les variable au travers de la d-séparation :

Théorème 2.1.1. *Étant donné un graphe \mathcal{G} , et P une distribution de probabilité fidèle aux indépendances conditionnelles que décrit \mathcal{G} , alors*

P modélise l'indépendance entre deux groupes de variables X et Y sachant Z si et seulement s'ils sont d-séparés étant donné Z .

Cette notion induit une manière de factoriser la distribution de probabilité associée aux variables d'un tel système.

Supposons que, dans un GOA \mathcal{G} , X_i a pour parent l'ensemble de noeud $\mathcal{P}(X_i)$. Suivant la caractérisation mentionnée ci-dessus, on a alors que X_i est d-séparé de ses non-descendants étant donné $\mathcal{P}(X_i)$. Si on numérote les variables aléatoires de telle sorte que les parents de X_i sont dans l'ensemble X_1, \dots, X_{i-1} (ce qui est possible de faire puisque le graphe ne contient pas de cycle), et si θ_i est une mesure de probabilité associée à X_i qui tient compte l'indépendance conditionnelle entre les variables, on aurait alors, pour tout i ,

$$\theta_i = \mathbb{P}(X_i \mid X_1, \dots, X_{i-1}) = \mathbb{P}(X_i \mid \mathcal{P}(X_i)).$$

Ce qui permet de réécrire l'équation (2.1)

$$\mathbb{P}(X_1, \dots, X_n) = \prod_i \theta_i = \mathbb{P}(X_1) \prod_i \mathbb{P}(X_i \mid \mathcal{P}(X_i)).$$

La factorisation précédente mène à la définition de la compatibilité de Markov qui permettra d'énoncer un théorème résultant de la d-séparation.

Définition 2.1.5 (Compatibilité de Markov). *Soit $\mathcal{G} = (\mathcal{X}, \mathcal{E})$ un graphe orienté acyclique, X_1, X_2, \dots, X_k un ensemble de variables aléatoires représenté par les sommets du GOA et P une distribution de probabilité de densité/fonction de masse f et $\mathcal{P}(X_i)$ l'ensemble des parents de X_i . Si P se factorise sur \mathcal{G} , c'est-à-dire que f admet une factorisation*

$$f_{X_1, X_2, \dots, X_k}(x_1, x_2, \dots, x_k) = \prod_{i=1}^k f_{X_i \mid \mathcal{P}(X_i)}(x_i \mid \mathcal{P}(x_i)),$$

*alors on dit que \mathcal{G} et P sont **compatibles**.*

Cette notion de compatibilité permet d'énoncer le theoreme qui lie la d-séparation et l'indépendance conditionnelle.

Théorème 2.1.2. *Soit $\mathcal{G} = (\mathcal{X}, \mathcal{E})$ un GOA et P une distribution de probabilité compatible avec \mathcal{G} . Alors, si deux ensembles disjoints X et Y sont d-séparés étant donné Z , il s'ensuit que*

$$\mathbb{P}(X, Y \mid Z) = \mathbb{P}(X \mid Z) \mathbb{P}(Y \mid Z).$$

Voici maintenant une définition formelle des réseaux bayésiens.

Définition 2.1.6 (Réseaux bayésiens). *Soit $\mathcal{G} = (\mathcal{X}, \mathcal{E})$ un GOA où $\mathcal{X} = \{X_1, \dots, X_n\}$ est un ensemble de variables aléatoires associées aux différents sommets du graphe et \mathcal{E} représente les arrêtes. Soit P une distribution de probabilité compatible avec \mathcal{G} . Alors, la paire (\mathcal{G}, P) est un **réseau bayésien**.*

Un réseau bayésien offre donc une approche pour la construction de distributions de probabilité, par la spécification des indépendances locales entres les variables, puis par l'élicitation des distributions de probabilités associées à chacun des termes $f_X(X_i \mid \mathcal{P}(X_i))$ de la factorisation.

2.2 Diagnostique de normalité

Tel que mentionné au tout début du chapitre, l'estimateur de densité à noyau a été utilisé pour estimer la densité de la charge virale. Toutefois, avant d'utiliser l'estimation non-paramétrique il a été nécessaire de réfuter l'hypothèse de log-normalité de la distribution empirique des données à l'aide de tests de normalité.

2.2.1 Diagramme quantile-quantile

Le diagramme quantile-quantile (QQ plot) est un outil de comparaison graphique. Il permet de comparer la distribution empirique de deux jeux de données ou encore de

comparer la distribution des données à une distribution paramétrique théorique dite de référence. Dans le cas de la loi normale, on parle alors de test de la droite de Henry [26, 45]. Dans le cas de la présente recherche, ce test a été utilisé afin de vérifier l'inadéquation de la distribution des données de charge virale des donneurs infectés à la loi log-normale. Afin d'effectuer le test, tel que décrit dans le livre *Probabilités, analyses des données et statistiques* [27], on doit

1. Ordonner les données x_i par ordre croissant.
2. Calculer les $F_i = \frac{i-3/8}{n+1/4}$, soit la fonction de répartition empirique associée à chaque x_i .
3. Calculer les quantiles z_i d'ordre F_i et dénormaliser les données : $y_i = z_i \times s + \bar{x}$ où \bar{x} et s sont respectivement la moyenne et l'écart-type des données.
4. Tracer les couples (x_i, y_i) .

Si le nuage de points est aligné sur la droite de référence, on dit qu'il forme une droite de Henry et il est possible de conclure que la distribution des données s'apparente à une distribution normale. Toutefois, lorsque le nuage de points forme plutôt une courbe, on peut conclure que les données sont probablement asymétrique. Une tendance à suivre la droite au centre et s'en éloigner aux extrémités traduit probablement la présence de valeurs extrêmes dans les données.

Puisqu'il s'agit d'un outil de diagnostic visuel, il est nécessaire de confirmer, à l'aide des coefficients d'asymétrie et d'aplatissement et d'un test statistique présenté à la prochaine section, l'hypothèse de normalité déduite à partir de l'allure de la droite de Henry.

2.2.2 Test de Shapiro-Wilk

Le test de Shapiro-Wilk [29] est un test très répandu. Il a été choisi ici puisqu'une comparaison de la puissance entre plusieurs tests de normalité a démontré qu'il s'agit d'un des tests les plus puissants pour des échantillons de taille plus grande que 50 [28]. La statistique du test prend des valeurs entre 0 et 1 s'écrit comme suit :

$$W = \frac{\left(\sum_{i=1}^n a_i y_{(i)}\right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

où $y_{(i)}$ est la statistique du i^e ordre, \bar{y} est la moyenne échantillonnale et $(a_1, \dots, a_n) = \frac{m^T v^{-1}}{\sqrt{m^T v^{-1} v^{-1} m^T}}$ avec $m^T = (m_1, \dots, m_n)$ un vecteur des valeurs attendues des statistiques d'ordre d'une loi normale standard et $v = \{v_{ij}\}$ la matrice $n \times n$ de covariance associée à m .

Lorsque $W = 1$, les données sont de distribution normale et plus la valeur de W s'approche de 0, plus les données s'éloignent d'une distribution normale. Pour de petits échantillons (< 50) il est possible de calculer W et de se référer à une table de Shapiro-Wilk afin de déterminer la valeur du seuil de rejet de l'hypothèse nulle selon laquelle l'échantillon suit une loi normale. Plus généralement, pour des échantillons de taille $3 \leq n \leq 5000$, un algorithme est utilisé. Si la valeur p est significative, généralement $p < 0,05$ lorsque le seuil de confiance est à 95%, on rejette l'hypothèse nulle.

Si l'hypothèse de normalité est rejetée et que la distribution empirique des données ne s'apparente à aucune autre loi paramétrique, une approche couramment employée consiste en utiliser une technique d'estimation non paramétrique de la densité. Dans le cas de l'estimation de la charge virale, l'hypothèse de log-normalité est réfutée et on

utilisera l'estimation à noyau afin d'estimer la distribution des données.

2.3 Estimation non paramétrique de la densité

L'estimation non paramétrique d'une densité [49] s'avère très utile lorsque l'on cherche à simuler des valeurs tout en restant le plus fidèle possible aux données observées, que l'on suppose être des réalisations de variables aléatoires i.i.d. et lorsqu'aucune distribution paramétrique ne semble s'ajuster à la distribution des données observées. Les estimateurs à noyaux est l'un des types d'estimation non paramétrique de la densité.

Estimateurs à noyaux

L'estimation de densité par noyaux est utilisée très fréquemment. Elle se base sur les données échantillonales et permet d'obtenir une fonction de densité plutôt lisse et continue qui est fidèle à la distribution des données. Dans le cas présent, la simulation d'une charge virale associée à un don de plasma VIH positif a nécessité le recours à l'estimateur de densité à noyau. La procédure repose sur le choix d'un noyau K et d'un paramètre de lissage $h > 0$.

Définition 2.3.1 (Estimateur à noyaux). *Soit un ensemble $\mathcal{X} = \{X_1, \dots, X_n\}$ d'observations admettant une densité f . L'estimateur à noyaux est tel que la densité à un point x quelconque est donnée par*

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

où K , le noyau, est une fonction de $\mathbb{R} \rightarrow \mathbb{R}$ telle que $\int_{\mathbb{R}} K(u) du = 1$, $K(u) \geq 0$ et $K(u) = K(-u)$. h est le paramètre de lissage.

Parmi les noyaux fréquemment utilisés en pratique, on retrouve le noyau triangulaire $K(u) = \frac{1}{2}\mathbb{1}_{[0,1]}(u)$, le noyau d'Epanechnikov $K(u) = \frac{3}{4}(1-u^2)\mathbb{1}_{[0,1]}(u)$ et le noyau gaussien

$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$. Pour la suite, nous considérerons uniquement le noyau gaussien. L'estimateur de densité à noyau gaussien est donc donné par l'expression

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{x - X_i}{h}\right)^2\right].$$

Le choix d'un bon paramètre de lissage est davantage important que le choix du noyau tant que celui-ci respecte les conditions énoncées précédemment.

2.3.1 Choix du paramètre de lissage

Le choix du paramètre de lissage influence grandement l'allure de la courbe de densité. D'abord, voyons l'effet de la valeur de h . Lorsque h est petit, c'est-à-dire qu'un moins grand nombre de valeurs sont considérées dans l'approximation locale, cela entraîne une plus grande variance, mais un plus petit biais. Graphiquement, cela donne lieu à une courbe beaucoup moins lisse, on parle alors de sous-lissage. À l'opposé, lorsque h est très grand, un plus grand nombre de données est utilisé dans l'estimation, ce qui donne lieu à une courbe lisse qui risque de mal représenter les particularités liées aux données. Le sur-lissage entraîne une petite variance, mais un grand biais. On peut très bien voir, sur la figure suivante, l'effet de la valeur de h . Cette figure est obtenue en estimant la densité de la charge virale basé sur les valeurs retrouvées dans une base de donnée utilisée dans la présente recherche et contenant des valeurs de charge virale associé à des dons sanguins ayant testé positif au VIH. Trois estimations ont été effectuées avec comme seule modification, la valeur du paramètre de lissage.

Sur la figure 2.2, des valeurs arbitraires ont été assignées pour les valeurs extrêmes de h . La valeur $h = 8628$ représente celle issue de la règle empirique de Silverman [46]. On remarque que, lorsque h est petit (courbe rouge), il y a beaucoup de bruit dans la courbe

d'estimation de la densité. À l'opposé, lorsque h est grand (courbe verte), la courbe est très lisse et l'on risque d'obtenir une estimation moins fidèle à la densité des données.

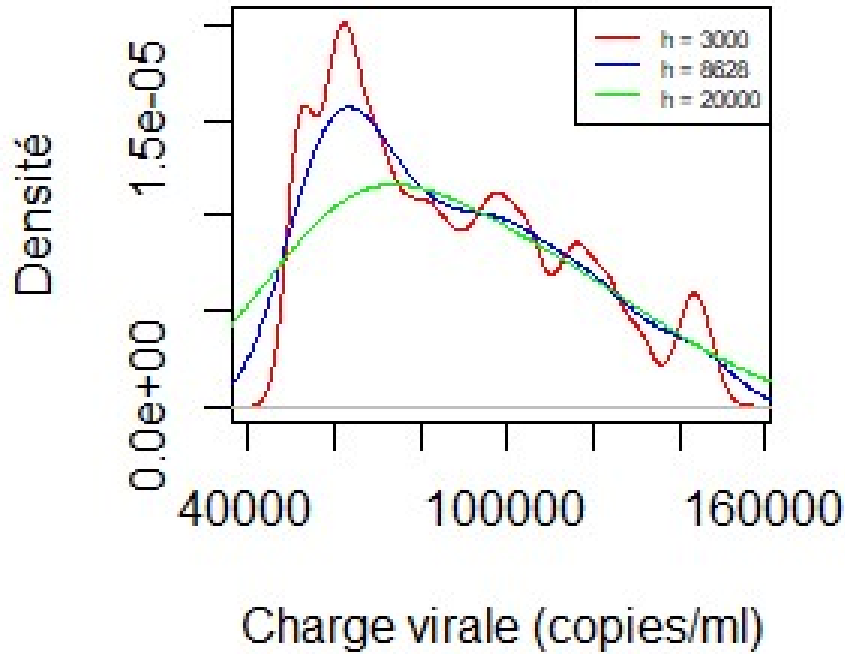


FIGURE 2.2 – Densité de la charge virale.

Il existe plusieurs techniques pour déterminer la valeur optimale pour le paramètre de lissage. Dans la présente recherche, trois techniques différentes ont été considérées, soit la sélection du paramètre par validation croisée [47, 49], par la règle de Silverman [46] ("Rule of Thumb") (ROT) et la méthode de Sheather et Jones [25]. Pour sélectionner le paramètre de lissage optimal, on cherche à minimiser l'erreur quadratique intégrée (EQI)

$$EQI(\hat{f}) = \int_{\mathbb{R}} (\hat{f}(x, h) - f(x))^2 dx,$$

où \hat{f} est l'estimateur à noyau tel que défini précédemment. Notez ici que $(\hat{f}(x, h))$ est

aléatoire comme fonction de X_1, \dots, X_n , alors $EQI(\hat{f})$ est aléatoire. Voyons maintenant les trois méthodes citées en détail.

1. *Validation croisée* [47, 49]

Le choix du paramètre de lissage par validation croisée est une technique très répandue. L'idée est de déterminer h en minimisant l'erreur quadratique moyenne intégrée (EQMI) qui est directement liée à l'erreur quadratique moyenne (EQM) et qui se définit comme suit :

$$EQMI(\hat{f}) = \mathbb{E} \left[\int_{\mathbb{R}} (\hat{f}(x, h) - f(x))^2 dx \right].$$

En développant le carré, on obtient

$$\begin{aligned} EQMI(\hat{f}) &= \mathbb{E} \left[\int_{\mathbb{R}} (\hat{f}(x, h)^2 - 2\hat{f}(x, h)f(x) + f(x)^2) dx \right] \\ &= \mathbb{E} \left[\int_{\mathbb{R}} (\hat{f}(x, h)^2) dx \right] - 2\mathbb{E} \left[\int_{\mathbb{R}} (\hat{f}(x, h)f(x)) dx \right] + \mathbb{E} \left[\int_{\mathbb{R}} (f(x)^2) dx \right] \\ &= \mathbb{E} \left[\int_{\mathbb{R}} (\hat{f}(x, h)^2) dx \right] - 2\mathbb{E} \left[\int_{\mathbb{R}} (\hat{f}(x, h)f(x)) dx \right] + \int_{\mathbb{R}} (f(x)^2) dx. \end{aligned}$$

On cherche donc à minimiser l'équation précédente par rapport à h . Il est donc équivalent de minimiser

$$\mathbb{E} \left[\int_{\mathbb{R}} (\hat{f}(x, h)^2) dx \right] - 2\mathbb{E} \left[\int_{\mathbb{R}} (\hat{f}(x, h)f(x)) dx \right], \quad (2.2)$$

puisque $f(x)$ ne dépend pas de h . Il est possible d'estimer (2.2) par

$$CV(h) = \int_{\mathbb{R}} (\hat{f}(x, h)^2) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}^{-i}(X_i, h),$$

où

$$\hat{f}^{-i}(X_i, h) = \frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{x_i - X_j}{h}\right)$$

est l'estimateur obtenu des données auxquelles on a retirée la i -ème observation.

Finalement, on sélectionne le paramètre de lissage en minimisant $CV(h)$ par rapport à h :

$$\hat{h}_{CV} = \operatorname{argmin}_{h>0} CV(h).$$

2. Règle de Silverman [46]

L'idée est encore de minimiser l'EQMI, mais en supposant que f est gaussienne. On a donc

$$\begin{aligned} EQMI(\hat{f}) &= \mathbb{E} \left[\int_{\mathbb{R}} (\hat{f}(x, h) - f(x))^2 dx \right] \\ &= \int_{\mathbb{R}} \text{Biais}^2(\hat{f}(x)) dx + \int_{\mathbb{R}} \text{Var}(\hat{f}(x)) dx \\ &= \frac{h^4}{4} \mu_2^2(K) R(f'') + \frac{1}{nh} R(K) + O(h^4) + O\left(\frac{4}{n}\right), \end{aligned}$$

où $\mu_2(K) = \int_{\mathbb{R}} z^2 K(z) dz$, $R(f'') = \int_{\mathbb{R}} (f''(x))^2 dx$ et $R(K) = \int_{\mathbb{R}} (K(z))^2 dz$.

On considère seulement l'approximation asymptotique de l'erreur quadratique moyenne intégrée (EQMIA), c'est-à-dire

$$EQMIA(\hat{f}) = \frac{h^4}{4} \mu_2^2(K) R(f'') + \frac{1}{nh} R(K).$$

Le paramètre de lissage qui minimise l'EQMIA est donc

$$h_{EQMIA} = \left(\frac{R(K)}{n \mu_2^2(K) R(f'')} \right)^{\frac{1}{5}}.$$

Tel que mentionné plus haut, pour l'évaluation du paramètre h à l'aide de la ROT de Silverman, on considère que $f(x)$ a la densité d'une loi normale $\mathcal{N}(\mu, \sigma^2)$ alors,

$$R(f'') = \frac{3\sigma^{-5}}{8\sqrt{\pi}}, \quad R(K) = \frac{1}{2\sqrt{\pi}} \quad \text{et} \quad \mu_2(K) = 1.$$

D'où l'on obtient le critère de sélection du paramètre de lissage par la règle de Silverman :

$$\hat{h}_S = \left(\frac{4}{3n} \right)^{\frac{1}{5}} \hat{\sigma},$$

où $\hat{\sigma} = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n) \right)^{\frac{1}{2}}$ est l'estimateur de l'écart-type de $f(x)$.

3. Méthode de Sheather et Jones

Reconsidérons le paramètre de lissage qui minimise l'EQMIA, c'est-à-dire

$$h_{EQMIA} = \left(\frac{R(K)}{n\mu_2^2(K)R(f'')} \right)^{\frac{1}{5}}.$$

Comme précédemment, il s'agit de trouver le meilleur estimateur possible pour $R(f'')$ avec R tel que $R(g) = \int (g(x))^2 dx$. En fait, $R(f'')$ est une mesure de l'ondulation de la courbe. Si la courbe est très ondulée, il y a beaucoup de bruit et il est préférable de réduire la quantité de données utilisées pour l'estimation. Il s'agit en fait d'utiliser un paramètre de lissage h plus petit. Dans le cas contraire, si la courbe est très peu ondulée, on veut ajouter du bruit, c'est-à-dire que l'on considère davantage de données dans l'estimation, ce qui mène à l'utilisation d'un h plus grand.

L'idée derrière la méthode de Sheather et Jones est d'inclure un biais dans l'estimation de $R(f'')$, ce qui mène à l'estimation de $R(f^{(3)})$. On estime d'abord $R(f'')$ par $R(\hat{f}'')$ avec \hat{f}'' un estimateur de f'' qui est déterminé à l'aide d'un choix de paramètre de lissage sous-optimal α . On estime $R(f'')$ par

$$S(\alpha) = \frac{1}{n(n-1)\alpha^5} \sum_i \sum_j L^{(4)} \left(\frac{x_i - x_j}{\alpha} \right),$$

où $L^{(4)}$ est la 4^{ième} dérivée du noyau L .

Dans l'article de Sheather et Jones (1991) [25], on définit le paramètre de lissage sous-optimal comme suit

$$\alpha = \frac{D(L)}{(nR(\hat{f}^{(3)}))^{\frac{1}{7}}},$$

avec

$$D(L) = \left(\frac{2L^{(4)}(0)}{\mu_2^2} \right)^{\frac{1}{7}}.$$

Le paramètre de lissage optimal de Sheather et Jones h_{SJ} est la solution de l'équation

$$\left(\frac{R(K)}{n\mu_2^2 S(\alpha)} \right)^{\frac{1}{5}} - h = 0. \quad (2.3)$$

Pour le noyau gaussien, on substitue $L^{(4)}$ par $\phi^{(4)}$ avec ϕ la densité d'une loi normale standard. L'estimateur h_{SJ} suivant l'utilisation d'un noyau gaussien est la solution de l'équation (2.3) avec

$$S(\alpha) = \frac{1}{n(n-1)\alpha^5} \sum_i \sum_j \phi^{(4)}\left(\frac{x_i - x_j}{\alpha}\right)$$

et tel que

$$\alpha = 1,357 \left(\frac{S(a)}{T(b)} \right)^{\frac{1}{7}} h^{\frac{5}{7}}, \quad T(b) = \frac{-1}{n(n-1)b^7} \sum_i \sum_j \phi^{(6)}\left(\frac{x_i - x_j}{b}\right)$$

avec

$$a = \frac{0,920\hat{\lambda}}{n^{\frac{1}{7}}} \quad \text{et} \quad b = \frac{0,912\hat{\lambda}}{n^{\frac{1}{9}}},$$

où $\hat{\lambda}$ est l'intervalle interquartile de l'échantillon ou encore toute autre mesure d'échelle associée à l'échantillon.

L'estimation de la densité avec l'estimateur à noyau est performante pour définir la densité de données à valeur dans \mathbb{R} . Toutefois, lorsque le support est borné ou semi-borné,

par exemple $(0, \infty)$, comme dans le cas de la charge virale, l'estimation de \hat{f}_x pourrait admettre une probabilité positive à des valeurs négatives proches de 0 et donc ne pas représenter adéquatement la distribution des données observées. Dans ce cas, il est possible d'appliquer une transformation aux données initiales afin de s'assurer de la cohérence du support.

Il est possible de choisir le paramètre de lissage avec les mêmes techniques lorsque l'on applique une transformation aux données. Par exemple, si on utilise la transformation logarithmique exposée dans la prochaine section, on a qu'à calculer h à partir de l'échantillon de données transformées soit $Y_1 = \log X_1, \dots, Y_n = \log X_n$.

2.3.2 Estimateur de densité à noyau log-normal

L'estimateur à noyau est très performant lorsque le support des données est l'axe réel. Toutefois, tel que discuté précédemment, lorsque le support est borné, par exemple dans le cas de la présente recherche où les valeurs de charge virale sont uniquement positives et dont les valeurs peuvent être près de zéro, l'estimateur peut ne pas être adéquat car il peut permettre des poids de densité positive à l'extérieur du support, et ce, surtout en présence de valeurs près de zéro. Il est donc pertinent d'effectuer une transformation sur les données afin de s'assurer de la cohérence de l'estimation. Lorsque le support est $(0, +\infty)$, il est approprié d'effectuer une transformation logarithmique.

Soit donc X une variable aléatoire positive. On pose alors $Y = \log X$. Il est possible

d'obtenir la densité de X en fonction de celle de Y :

$$\begin{aligned} f_X(x) &= f_Y(\log x) \frac{d}{dx}(\log x) \\ &= \frac{f_Y(\log x)}{x} \end{aligned}$$

pour $x > 0$.

Il est possible d'obtenir, suite à la transformation, une estimation de la densité de X avec l'estimateur à noyau.

$$\begin{aligned} \hat{f}_X(x) &= \frac{\hat{f}_Y(\log x)}{x} \\ &= \frac{1}{nhx} \sum_{i=1}^n K\left(\frac{\log x - \log X_i}{h}\right). \end{aligned}$$

Finalement, dans le contexte où l'on utilise un noyau gaussien, appliquer une transformation logarithmique aux données équivaut à utiliser un noyau log-normal. L'estimateur devient donc :

$$\begin{aligned} \hat{f}_X(x) &= \frac{1}{nhx} \sum_{i=1}^n \phi\left(\frac{\log x - \log X_i}{h}\right) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{hx\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log x - \log X_i}{h}\right)^2\right]. \end{aligned}$$

Toutefois, cette technique d'estimation de la densité donne de mauvais résultats à la frontière sauf dans le cas où $f_X(0) = 0$. Dans le cas de l'utilisation d'un noyau gaussien, on a que $\lim_{x \rightarrow 0^+} \frac{\hat{f}_Y(y)}{x}$ est de forme indéterminée puisque $\hat{f}_Y(y)$ et $\frac{1}{x}$ tendent respectivement vers $-\infty$ et vers ∞ lorsque $x \rightarrow 0^+$. Cela pose donc un problème près de la borne inférieure. Geenens et Wang (2018) proposent plutôt une méthode basée sur une estimation de la fonction près de zéro à l'aide de l'expansion de Taylor (local likelihood density estimation) afin de pallier aux possibles comportement erratiques de la fonction près de 0

[15]. C'est donc une fonction basé sur leur méthode qui à été utilisée dans les simulations.

Les sections suivantes présentent l'exemple d'un réseau bayésien, il s'agit en fait du réseau construit dans le cadre de cette recherche.

2.4 Évaluation du risque

Il est possible d'évaluer le risque, à partir d'un réseau bayésien, avec deux techniques différentes. On peut simuler les éléments d'une population représentative de la réalité un à un et de se baser sur les résultats obtenus pour estimer le risque (simulation ligne par ligne) ou encore estimer le risque directement à partir de la distribution de probabilités jointes. Dans le présent cas, la simulation ligne par ligne a été utilisée.

Ce type d'évaluation du risque consiste à simuler chacun des éléments d'une population à partir des distributions de probabilités associées à chacune des caractéristiques pouvant être définies chez un même individu, tout en considérant les dépendances possibles entre chacune de ces caractéristiques. Le nombre d'individus simulés est déterminé à l'avance. À partir de la population simulée, il est possible d'évaluer l'occurrence d'un certain événement, donc le risque lié à cet événement. On utilise ensuite une procédure de type Monte Carlo c'est à dire que l'on refait l'exercice précédent, de manière indépendante, un très grand nombre de fois puis on calcule l'espérance de tous les résultats obtenus. En fait, on utilise la distribution de probabilité jointe dans le but d'estimer la probabilité marginale. La simulation de type Monte Carlo est en fait une technique d'estimation dans laquelle on génère des variables aléatoires à partir de distributions de probabilité. Elle permet

d'obtenir une estimation de l'incertitude du modèle. Afin de bien estimer les distributions de probabilités associées aux différents paramètres du modèle, il convient de tester la correspondance de la distribution à une loi paramétrique. Dans le cas où la distribution des données est plutôt atypique, on effectue une estimation non paramétrique de la densité.

Dans le contexte de l'évaluation du risque d'obtenir un lot de plasma contaminé au VIH, on simule individuellement et indépendamment chacun de ces dons de plasma se retrouvant dans le lot. Pour chacun des dons, on détermine certaines caractéristiques associées au donneur. Par exemple, on détermine le sexe et l'âge du donneur ayant effectué le don puisque la prévalence et l'incidence varient, entre autres, en fonction de ces facteurs. Ces caractéristiques sont attribuées à partir de proportions estimées dans la population de référence et auxquelles une distribution a priori a été associée afin d'introduire une variation permettant de pallier à l'incertitude liée aux données.

2.4.1 Exemple

Tel que discuté dans le premier chapitre, on tente d'estimer le risque résiduel d'obtenir un lot de plasma destiné au fractionnement soit infecté par le VIH. Les dons proviennent de dons de plasma récoltés par aphérèse suite à une modification du critère d'admissibilité pour les HARSAH. Afin de déterminer la probabilité d'obtenir un lot infecté, il faut d'abord évaluer la probabilité qu'un don soit infecté par le VIH. On identifie chacun des facteurs qui ont une influence sur le risque. On modélise les dépendances entre ces facteurs à l'aide d'un GOA ce qui permettra d'obtenir plus facilement une formule pour le calcul de la probabilité qu'un don de plasma soit infecté par le VIH. Le GOA de la figure 2.3 représente les différents éléments qui influent le risque d'infection d'un lot de plasma

de fractionnement. Ces éléments sont les suivants :

1. Âge : Probabilité de se retrouver dans un des trois groupes d'âge suivants 17/18 à 24 ans, 25 à 44 ans et 45 ans et plus.
2. Sexe : Probabilité que le don provienne d'un donneur masculin ou féminin.
3. HARSAH : Probabilité que le don provienne d'un donneur HARSAH.
4. Statut du donneur : Probabilité que le don soit effectué par un nouveau donneur (ou un donneur connu).
5. Compliance : Probabilité que le donneur ne soit pas compliant au critère HARSAH.
6. Prévalence : Probabilité que le don provienne d'un donneur présentant une infection prévalente du VIH.
7. Incidence : Probabilité que le don provienne d'un donneur présentant une infection incidente du VIH.
8. Période muette (silencieuse)(PS) : Probabilité que le don ait été effectué durant la période silencieuse.

En utilisant ces probabilités, on obtient la probabilité qu'un don infecté au VIH ou plutôt la distribution de probabilité conjointe pour ces variables :

$$\mathbb{P}(\hat{age}, sexe, HARSAH, statut, compliance, prévalence, incidence, PS). \quad (2.4)$$

Cette dernière se factorise comme suit :

$$\begin{aligned} (2.4) &= \mathbb{P}(\hat{age}) \mathbb{P}(sexe \mid \hat{age}) \mathbb{P}(HARSAH \mid sexe) \mathbb{P}(statut \mid \hat{age}, sexe, HARSAH) \\ &\quad \times \mathbb{P}(compliance \mid HARSAH) \mathbb{P}(prévalence \mid sexe, HARSAH, statut, compliance) \\ &\quad \times \mathbb{P}(incidence \mid sexe, HARSAH, statut, compliance, prevalence) \mathbb{P}(PS \mid incidence). \end{aligned}$$

Cette décomposition est basé sur une concertation avec plusieurs experts internationaux. Ces derniers ont participé au choix des paramètres ainsi qu'à l'identification des dépendances entre ces derniers. Les probabilités conditionnelles seront utilisées afin de simuler individuellement chacun des dons de plasma se retrouvant dans un bassin de dons destiné au fractionnement. Le processus de simulation sera expliqué plus en détail au prochain chapitre.

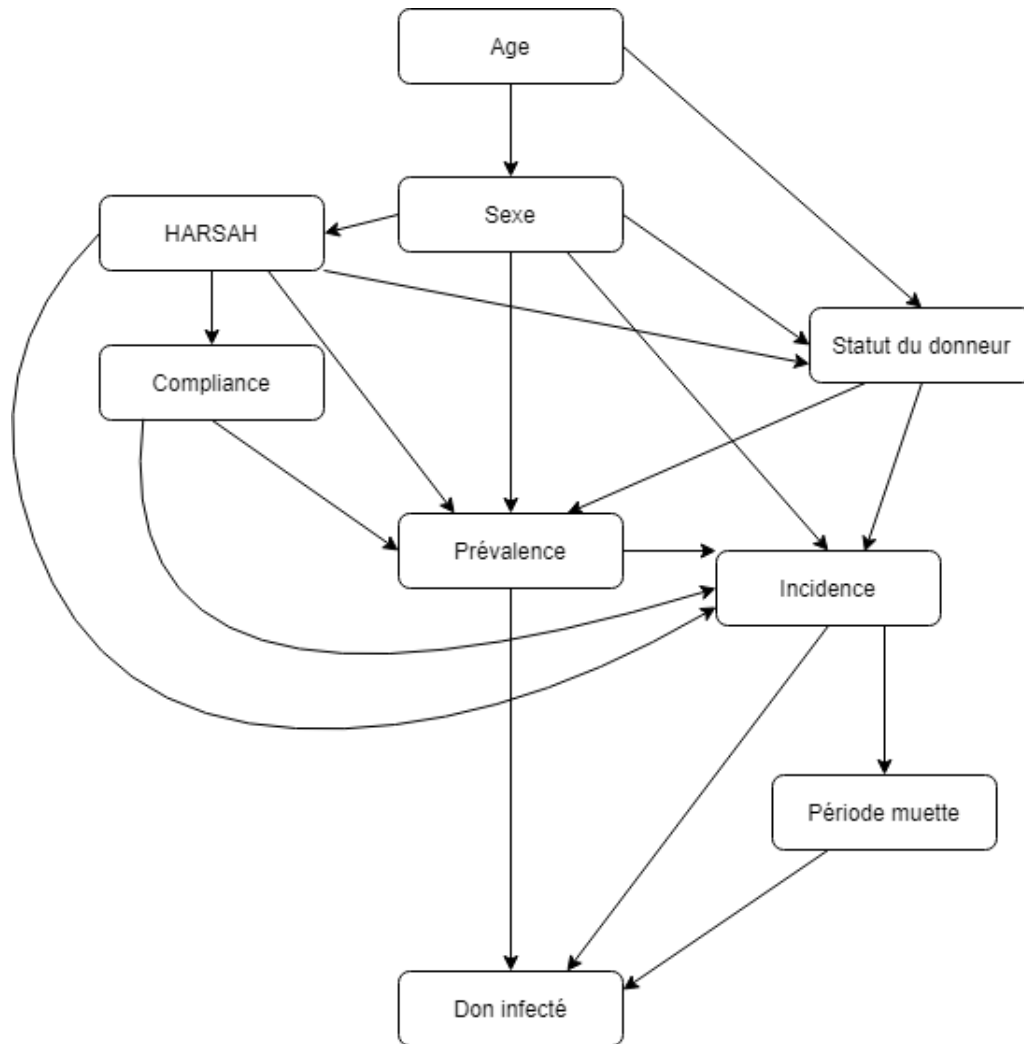


FIGURE 2.3 – GOA associé au risque d'infection d'un don.

2.4.2 Probabilités conditionnelles

Une loi *a priori* de type beta pour la distribution de certains paramètres de la modélisation a été attribuée afin de modéliser la variabilité associée à l'incertitude entourant la connaissance de ces derniers. Une fois que les paramètres de la loi beta sont choisis, il est possible de déterminer, par simulation, si un don possède une caractéristique ou non. La probabilité peut être simulée de deux méthodes différentes.

1. La première méthode consiste à échantillonner un nombre (probabilité) à partir de la distribution beta associée au paramètre. Par exemple, afin de déterminer si le don simulé provient d'un donneur HARSAH, on détermine la probabilité qu'un don provienne d'un donneur HARSAH ($\mathbb{P}(HARSAH \mid Sexe)$) en se basant sur la distribution beta associée à ce paramètre. Une fois que l'on a estimé la probabilité, on simule une Bernoulli en fonction de cette probabilité.
2. Dans la deuxième méthode, on simule une Bernoulli en fonction des paramètres de la loi beta plutôt qu'à partir des probabilités, comme suit : Soit $Y|P \sim Ber(P)$ et $P \sim B(\alpha, \beta)$. On a,

$$\begin{aligned}\mathbb{P}(Y = 1) &= \int_0^1 \mathbb{P}(Y = 1|P = p) f_{\alpha, \beta}(p) dp \\ &= \int_0^1 p f_{\alpha, \beta}(p) dp \\ &= \frac{\alpha}{\alpha + \beta}.\end{aligned}$$

Plus concrètement, si on réutilise l'exemple précédent, on a $Y = HARSAH \mid Sexe$, $1 = oui$ et $f_{\alpha, \beta}(p)$ est la distribution beta associée à la proportion de dons fait par des HARSAH.

L'idée derrière cette deuxième méthode est de possiblement réduire les variations dans les variables que l'on génère comparativement à la première méthode. Les deux méthodes ont

été comparées en testant leurs variances. Par exemple, si la variance associée aux résultats de la seconde méthode est significativement plus petite, cela implique que cette méthode induit moins de variation dans les variables générées. Les probabilités pour lesquelles il a été possible de simuler les deux méthodes sont $\mathbb{P}(HARSAH \mid \text{sexe})$, $\mathbb{P}(\text{Statut} \mid \text{âge, sexe, HARSAH})$ et $\mathbb{P}(\text{compliance} \mid HARSAH)$.

2.4.3 Test d'égalité des variances

Pour tester l'égalité des variances à l'aide de la statistique F , on doit d'abord vérifier la normalité des données.

1. *Test du khi-deux* [48].

Soit un échantillon de taille n . Afin de tester si les données $X = \{x_1, x_2, \dots, x_n\}$ suivent une loi normale, on découpe l'échantillon en k intervalles. On doit calculer les fréquences théoriques attendues à partir des données échantillonnales, la moyenne et l'écart-type dans ce cas-ci. La fréquence théorique attendue pour l'intervalle i est $E_i = np_i$ où p_i est la probabilité théorique. On doit s'assurer que, pour 80% des intervalles, on ait $E_i \geq 5$. On cherche donc à tester l'hypothèse :

$$H_0 : X \sim N(\mu, \sigma^2)$$

contre l'hypothèse :

$$H_1 : X \text{ ne suit pas une loi normale.}$$

Soit la statistique

$$\chi^2 = \sum_{i=1}^k \frac{(x_i - E_i)^2}{E_i}.$$

On rejette H_0 si $\chi^2 \geq \chi_\alpha^2(K - l - 1)$ où l = nombre de paramètres estimés et α est le seuil de significativité du test. Par exemple, on peut utiliser la moyenne échantillonnale et l'écart-type échantillonnal afin de générer une distribution normale théorique. La valeur de l serait donc égale à 2.

2. *Test d'égalité de deux variances provenant de deux lois normales indépendantes* [48].

Soient $X = \{x_1, x_2, \dots, x_n\}$ de loi $N(\mu_X, \sigma_X^2)$ et $Y = \{y_1, y_2, \dots, y_m\}$ de loi $N(\mu_Y, \sigma_Y^2)$. On veut confronter $H_0 : \sigma_X^2 = \sigma_Y^2$ contre différentes hypothèses H_1 ($\sigma_X^2 \neq \sigma_Y^2$, $\sigma_X^2 > \sigma_Y^2$, $\sigma_X^2 < \sigma_Y^2$). Soit la statistique $F = \frac{S_X^2}{S_Y^2}$, avec $S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ et $S_Y^2 = \frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2$ et les différentes régions critiques associées à chacune des hypothèses alternatives :

H_0	H_1	Région critique
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 > \sigma_Y^2$	$F \geq F_\alpha(n-1, m-1)$
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 < \sigma_Y^2$	$\frac{1}{F} \geq F_\alpha(n-1, m-1)$
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 \neq \sigma_Y^2$	$F \geq F_{\frac{\alpha}{2}}(n-1, m-1)$ ou $\frac{1}{F} \geq F_{\frac{\alpha}{2}}(n-1, m-1)$

Pour l'hypothèse $H_1 : \sigma_X^2 > \sigma_Y^2$ on rejette H_0 au seuil α si $F < F_\alpha(n-1, m-1)$, où $F_\alpha(a, b)$ est le quantile d'ordre α d'une loi de Fisher de paramètres a et b . Il s'agit du cas d'intérêt puisque l'on cherche à tester si la deuxième méthode, présentée à la section 2.4.2, réduit les variations dans les variables générées. Si c'est le cas, la variance associée aux résultats obtenus à l'aide de la deuxième méthode devrait être inférieure à celle associée aux résultats obtenus avec la première méthode. On verra plus en détail, au chapitre trois, les résultats des tests ainsi que les conclusions que l'on peut en tirer.

CHAPITRE 3

Application

3.1 Introduction

Tel que mentionné précédemment, l'objectif de cette étude est d'évaluer le risque résiduel d'obtenir un lot de plasma de fractionnement infectieux suite à une modification du critère d'admissibilité des HARSAH au don de plasma récolté par aphérèse tout en modélisant l'effet de la TRP. L'aphérèse est un procédé selon lequel on récolte certains composants sanguins individuellement. Le fractionnement, quant à lui, consiste à séparer les différents composants du plasma afin de les transformer en produits stables et de les distribuer sous forme de médicaments.

Pour évaluer le risque, un modèle de simulation est construit. Ce dernier intègre des paramètres qui reflètent le mieux possible la réalité. Certaines des données nécessaires à la simulation sont très difficiles, voir impossibles à obtenir, entre autres, pour des raisons d'éthique ou de confidentialité. Il devient donc nécessaire d'estimer ces paramètres. Chacun des paramètres, leur valeur et distribution a priori associée sont déterminés en

collaboration avec un groupe d'experts, le groupe "Surveillance, Risk Assessment and Policy" (SRAP).

Ce chapitre présentera de façon détaillée chacune des parties du modèle. La simulation est divisée en quatre blocs :

1. Le profil du donneur ;
2. La probabilité d'infection ;
3. La charge virale ;
4. Le test TAN et la TRP.

3.2 Le modèle

À des fins de comparaison, deux hypothèses de politique d'exclusion HARSAH ont été simulées. La première hypothèse de simulation se base sur la politique d'exclusion canadienne actuelle, soit une exclusion du don pour trois mois suivant la dernière relation sexuelle HARSAH. Il est nécessaire d'effectuer une simulation pour ce critère d'exclusion puisque le changement de politique est récent (juin 2019) et que les données récoltées depuis ne sont pas suffisantes pour estimer le risque directement. Pour la deuxième hypothèse, le risque est estimé dans le contexte où il n'y aurait aucune exclusion liée à une relation HARSAH. En d'autres termes, les HARSAH sexuellement actifs seraient autorisés à donner du plasma par aphérèse. Pour chacune des hypothèses, quatre scénarios impliquant la variation de certains paramètres ont été simulés. Un scénario extrême a aussi été simulé pour l'hypothèse selon laquelle il n'y aurait aucune exclusion HARSAH. Ces derniers seront détaillés à la section 3.2.5.

Le premier bloc de simulation détermine les caractéristiques du donneur.

3.2.1 Profil du donneur

Chez les donneurs, les caractéristiques qui modifient la probabilité d'infection au VIH ont été identifiées. Pour simuler le profil du donneur, l'âge du donneur, son sexe, son appartenance à la communauté HARSAH ainsi que de sa conformité au critère d'exclusion HARSAH s'il est un homme, sont pris en compte. Ces caractéristiques ont été choisies en fonction des données disponibles et le choix a été validé par le SRAP.

Les données utilisées afin de calculer les proportions relatives au profil du donneur ont été fournies par Héma-Québec et la Société Canadienne du Sang. Ces données proviennent des informations récoltées, indépendamment par chacun des organismes, auprès des donneurs ayant fait des dons de plasma entre le 1er janvier 2018 au 31 décembre 2018 inclusivement. Pour chaque don de plasma effectué en 2018, les données incluent le sexe (homme ou femme), l'âge, le statut de donneur (nouveau ou connu), la date du prélèvement, et la durée, en jour, de l'intervalle interdon, soit le nombre de jour entre deux dons effectués par le même donneur. Comme le but est d'estimer le risque résiduel canadien d'obtenir un lot de plasma infecté, les données fournies par les deux organismes ont été mises en commun afin de calculer les différentes proportions nécessaires à la simulation.

Puisque la prévalence et l'incidence du VIH peuvent varier selon l'âge, le sexe et le statut du donneur, il a été décidé de d'abord diviser les données par groupe d'âge. À partir des données canadiennes de dons de plasma amassées en 2018, les proportions de dons ont été estimées pour chaque groupe. Une distribution a priori a été associée à chacune des proportions afin de bien modéliser la possible variabilité et l'incertitude associées aux

données.

Un fois les données divisées par groupe d'âge, les proportions d'hommes et de femmes ont été calculées pour chacune des catégories. Le tableau suivant présente les différentes proportions calculées pour la simulation.

Tableau 3.1 : Proportions

Variables			Estimation	
			Exclusion HARSAH de trois mois	Aucune exclusion HARSAH
Âge (proportion)	18-24		0.08269732 (0.0692859, 0.09640045)	
	25-44		0.2319182 (0.2160513, 0.2516827)	
	45+		0.6853845 (0.6579499, 0.705734)	
Proportion d'hommes par groupe d'âge	18-24		0.5824743 (0.5420628, 0.6362237)	0.5827171 (0.542311, 0.636455)
	25-44		0.6585760 (0.6168964, 0.7002833)	0.6588005 (0.6171323, 0.7004928)
	45+		0.7666982 (0.7480788, 0.80035796)	0.7668768 (0.7482669, 0.8005175)
Proportion de dons faits par de nouveaux donneurs par sexe et groupe d'âge	18-24	Homme	0.04479938 (0.0255102, 0.08035714)	
		Femme	0.05642973 (0.03461538, 0.08086253)	
	25-44	Homme	0.01781217 (0.01188707, 0.02282158)	
		Femme	0.03170097 (0.01619433, 0.04340568)	
	45+	Homme	0.004390975 (0.001818917, 0.007499464)	
		Femme	0.01235725 (0.006648936, 0.0206754)	

Afin de bien évaluer le risque lié à un changement de politique HARSAH, il est important d'estimer la proportion d'hommes HARSAH dans la population de donneurs ainsi que le taux de non-compliance au critère d'exclusion HARSAH chez les donneurs masculins. La proportion de donneurs HARSAH est basée sur une estimation de la proportion HARSAH

effectuée dans le cadre d’une politique d’exclusion de 12 mois [O’Brien et al., 2019] [11]. Cette valeur a été ajustée, comme le montre la prochaine formule, en fonction des deux hypothèses de politique d’exclusion. L’estimation de la proportion de donneurs HARSAH sous une exclusion de trois mois ($P_{\text{HARSAH 3-mois}}$) nécessite la connaissance de la proportion de donneurs HARSAH supplémentaires suivant une modification de politique de 12 mois à trois mois ($P_{\text{HARSAH supp 3-mois}}$). On a

$$P_{\text{HARSAH supp 3-mois}} = P_{\text{HARSAH actif 6 mois}} + P_{\text{HARSAH actif 12 mois pas 6 mois}} - P_{\text{HARSAH actif 3 mois}},$$

où $P_{\text{HARSAH actif 3 mois}}$ est le nombre de donneurs HARSAH sexuellement actifs au cours des trois derniers mois, $P_{\text{HARSAH actif 6 mois}}$ est le nombre de donneurs HARSAH sexuellement actifs au cours des six derniers mois et $P_{\text{HARSAH actif 12 mois pas 6 mois}}$ est le nombre de donneurs HARSAH sexuellement actifs au cours des six mois précédant les six derniers mois.

Il est donc possible d’estimer la proportion de donneurs HARSAH sous une hypothèse d’exclusion de trois mois. En effet, on obtient la proportion de donneur HARSAH après un changement de politique d’exclusion en faisant le ratio entre la somme de la proportion de donneurs HARSAH supplémentaires suivant un changement de politique et de la proportion de donneurs HARSAH avant le changement et la nouvelle proportion totale de donneurs soit $1 + P_{\text{HARSAH supp 3-mois}}$.

$$P_{\text{HARSAH 3-mois}} = \frac{P_{\text{HARSAH 12-mois}} + P_{\text{HARSAH supp 3-mois}}}{1 + P_{\text{HARSAH supp 3-mois}}},$$

où $P_{\text{HARSAH 12-mois}}$ est la proportion de donneurs HARSAH sous une politique d’exclusion du don de 12 mois suivant la dernière relation sexuelle HARSAH.

Dans le contexte où les HARSAH sexuellement actifs sont admissibles au don de plasma par aphérèse, on calcule la proportion de donneurs HARSAH sensiblement avec la même méthode. On considère que la proportion totale de donneurs s'ajuste en ajoutant la proportion de donneurs supplémentaire suite au changement de politique, soit les HARSAH sexuellement actifs dans les trois derniers mois ($1 + P_{\text{HARSAH actif 3-mois}}$). La proportion de donneurs HARSAH sous aucune politique d'exclusion ($P_{\text{HARSAH sans exclusion}}$) est donc calculée comme suit :

$$P_{\text{HARSAH sans exclusion}} = \frac{P_{\text{HARSAH 3-mois}} + P_{\text{HARSAH actif 3-mois}}}{1 + P_{\text{HARSAH actif 3-mois}}},$$

Une fois la proportion de donneurs HARSAH estimée, il est important de déterminer le taux de non conformité au critère d'exclusion de trois mois ou encore la proportion de donneurs HARSAH présentant un plus grand risque d'infection au VIH sous l'hypothèse selon laquelle il n'y aurait pas de politique d'exclusion HARSAH. Afin d'estimer le taux de non-compliance au critère 3-mois chez les HARSAH ($P_{\text{non-compliance HARSAH}}$), le taux de non-compliance chez les hommes sous exclusion HARSAH 3-mois ($P_{\text{non-compliance hommes 3-mois}}$) présenté dans O'Brien et al., 2020 [12] est utilisé. De plus, on a que si un homme est non-compliant au critère d'exclusion HARSAH, ce dernier est nécessairement HARSAH. Il est donc possible de calculer le taux de non-compliance chez les HARSAH plutôt que chez les hommes :

$$P_{\text{non-compliance HARSAH}} = \frac{P_{\text{non-compliance hommes 3-mois}}}{P_{\text{donneurs HARSAH}}}.$$

La proportion d'hommes à comportement HARSAH à risque élevé ($P_{\text{risque élevé}}$) est estimée à partir de la proportion de d'hommes non-compliants au critère d'exclusion HARSAH de 3 mois. On suit la même logique de calcul que dans O'Brien et al., 2019 [11]. La proportion

de d'hommes non-compliants est doublée puisque les donneurs non-compliants avant la levée de l'interdiction continuerait à donner du plasma et le nombre de HARSAH à risque élevé nouvellement éligibles au don est estimé proportionnel au nombre de HARSAH non-compliants. La formule pour obtenir $P_{\text{risque élevé}}$ est la suivante :

$$P_{\text{risque élevé}} = \frac{P_{\text{non-compliance hommes 3-mois}} \times 2}{1 + P_{\text{non-compliance hommes 3-mois}}}.$$

Les proportions de donneurs HARSAH et le taux de non-compliance associés à chacune des hypothèses de politique d'exclusion sont présentés dans le tableau 3.2. À la figure 3.1, on présente la section du graphe orienté acyclique (GOA) associée au profil du donneur.

Tel que mentionné plus haut, dans l'intention de bien modéliser la variabilité des données, des distributions a priori beta ont été associées aux différentes proportions estimées. Par exemple, pour ajouter une variation à l'estimation de la proportion de donneurs masculins et féminins, les données ont été catégorisées par mois, puis la proportion a été calculée pour chacun des mois.

Tableau 3.2 : Proportions de donneurs HARSAH et taux de compliance

Proportion de HARSAH dans la population générale	0,029 (0,018-0,039)	
	Exclusion 3-mois	Aucune exclusion
Proportion de donneurs HARSAH	0.016274 (0.009992-0.025424)	0.017257 (0.010981-0.026398)
Taux de non-compliance au critère HARSAH chez les hommes	0.001 (0.0003-0.002)	-----
Proportion d'hommes HARSAH à risque élevé	-----	0.001998 (0.0005998-0.00399)
Proportion de nouveaux donneurs HARSAH par groupe d'âge		
18-24	0.052228339 (0.034832578, 0.084295769)	0.056173633 (0.03978341, 0.08638746)
25-44	0.027890186 (0.022546688, 0.032407872)	0.03324231 (0.02820768, 0.03749886)
45 et plus	0.015786406 (0.01346682, 0.018589767)	0.02183817 (0.01965266, 0.02447949)

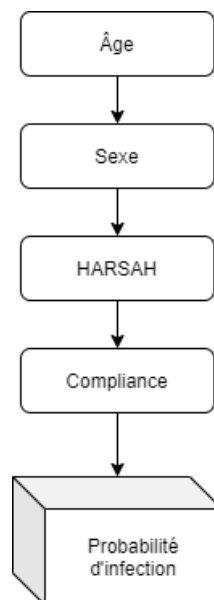


FIGURE 3.1 – GOA associé au profil du donneur.

Les bornes de l'intervalle de confiance à 95 % de la proportion calculée à partir des

données disponibles ont été utilisées pour estimer la distribution des paramètres dans le modèle. La valeur de la proportion calculée est associée à la médiane. Les valeurs associées à ces trois quantiles ont été utilisées pour déterminer les paramètres α et β de la distribution a priori beta. En effet, si on pose x_1, x_2 et x_3 les valeurs des proportions associées aux quantiles $p_1 = 0.025$, à la médiane et $p_3 = 0.975$ respectivement, on a

$$\begin{aligned} F(\alpha, \beta; x_i) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^{x_i} t^{\alpha-1}(1-t)^{\beta-1} dt \\ &= p_i. \end{aligned}$$

On réussit à estimer les paramètres α et β à l'aide de techniques numériques d'optimisation.

Comme discuté aux sections 2.3.5 et 2.3.6 du chapitre précédent, deux méthodes ont été utilisées pour simuler certaines des probabilités conditionnelles afin de déterminer si une technique mène à moins de variabilité dans les variables que l'on génère. Les simulations ont d'abord été effectuées à 30 reprises en utilisant chacune des deux méthodes. Une fois les données obtenues, un test du Khi-deux à été utilisé afin de vérifier la normalité des données pour chacun des échantillons de 30 valeurs. Par exemple, pour les données obtenues des simulations utilisant la deuxième méthode, on observe une variation entre 2,67 et 2,85. Les classes suivantes ont donc été choisies en respectant la condition $E_i \geq 5$ pour 80% des classes : $[2,67; 2,73), [2,73; 2,75), [2,75; 2,77), [2,77; 2,79), [2,79; 2,82), [2,82; 2,85]$. La moyenne échantillonnale ($\bar{x} = 2,75$) et l'écart-type échantillonnal ($s = 0,039$) ont été utilisés afin de générer une distribution normale théorique. Soit donc $K = 6$ le nombre de classes et $l = 2$ le nombre de paramètres utilisés dans l'estimation. Le degré de liberté (df) est égal à trois. On choisit un niveau de signification $\alpha = 0.05$. On obtient le tableau d'effectifs observés et théoriques suivant :

	[2, 67; 2, 73]	[2, 73; 2, 75]	[2, 75; 2, 77]	[2, 77; 2, 79]	[2, 79; 2, 82]	[2, 82; 2, 85]
Obs.	7	5	4	7	4	3
Prob.	0,202	0,171	0,201	0,183	0,171	0,065
Th.	6,06	5,13	6,04	5,50	5,13	1,95

Soit $H_0 : X \sim N(\bar{x}, s^2)$.

On rejette h_0 si $\chi^2 > \chi_{0,05}^2(3)$

Or, $\chi^2 = 2,06$ et $\chi_{0,05}^2(3) = 7,81$. On ne rejette donc pas H_0 au seuil $\alpha = 0,05$ et il est possible de supposer que les données ont une distribution normale.

Puisque la condition de normalité est respectée, on teste l'égalité des variances à l'aide de la statistique F . Dans le cas où les données n'auraient pas été normalement distribuées, on aurait pu avoir recours à un test non-paramétrique comme le test de Wilcoxon. On veut donc tester $H_0 : \sigma_X^2 = \sigma_Y^2$ contre $H_1 : \sigma_X^2 > \sigma_Y^2$. Les deux échantillons sont de taille $n = 30$, on a donc $df_X = df_Y = 29$ et on trouve $F = \frac{\sigma_X^2}{\sigma_Y^2} = \frac{0.001333}{0.001246} = 1,07$ et $F_{0,05}(29, 29) = 1,86$. On rejette H_0 si $F > F_{0,05}(29, 29)$. Avec une p-value de 0.4425, on ne rejette donc pas H_0 au seuil 0,05, on considère donc les variances des deux échantillons égales. Les conclusions sont les mêmes pour les résultats liés à la concentration. On ne peut donc pas conclure qu'une méthode est meilleure que l'autre. La deuxième méthode ne génère pas une plus faible variation dans les proportions simulées.

Lorsque les proportions sont estimées et qu'une distribution a priori leur est associée, elles sont utilisées afin de déterminer la probabilité d'infection d'un don.

3.2.2 Probabilité d'infection

L'objectif de ce bloc est de déterminer la probabilité d'infection d'un don en fonction des caractéristiques du donneur simulées au bloc précédent. Le GOA qui se rattache à ce bloc de simulation est le suivant :

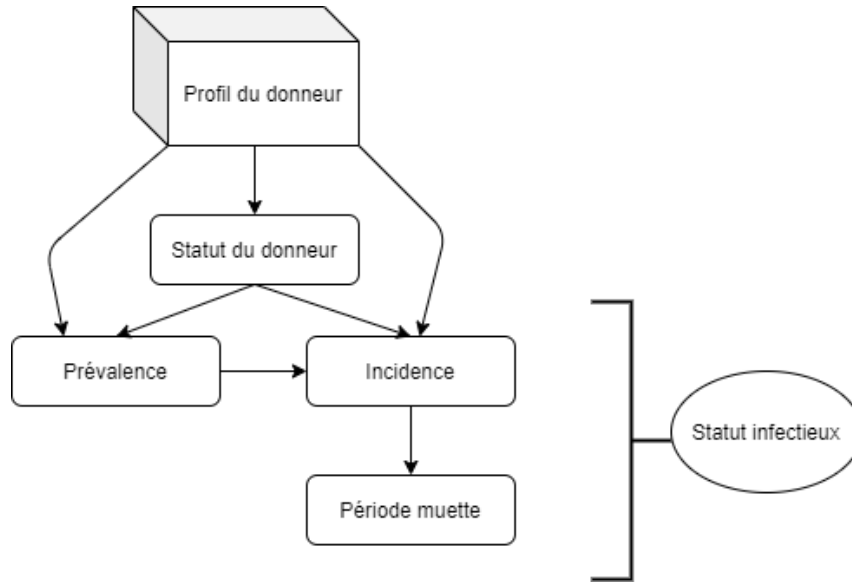


FIGURE 3.2 – GOA associé à la probabilité d'infection.

Le statut du donneur influence grandement le risque d'infectiosité. La première étape est de déterminer ce statut par simulation. Cette simulation est basée sur les proportions de nouveaux donneurs et de donneurs connus. Ces proportions sont estimées en fonction de l'âge et du sexe. La proportion de nouveaux donneurs HARSAH ne peut pas être estimée à partir de ces données. Il est important de considérer la proportion de nouveaux donneurs HARSAH qu'un changement de critère apporterait et de l'ajouter à la proportion de nouveaux donneurs estimée.

Afin de déterminer la probabilité d'infection d'un don au VIH, il faut connaître le taux

d'incidence du VIH chez les donneurs en fonction du sexe et du statut du donneur, soit la proportion de nouveaux cas de VIH dans chacune de ces populations. Il faut aussi détenir le taux de prévalence du VIH, soit la proportion d'individus atteints du VIH dans la population de donneurs. L'incidence et la prévalence doivent être estimées en fonction du sexe ainsi que du statut du donneur. Ces valeurs ont été obtenues à partir des données de collecte recueillies par Héma-Québec et la SCS sur 5 ans soit du 1^{er} juillet 2014 au 30 juin 2019. L'incidence chez les nouveaux donneurs a été obtenue en multipliant l'incidence estimée chez les donneurs connus par un facteur de correction de 1,65 ($\pm 1,0334$) variant selon une distribution normale tronquée à 0. Ce facteur de correction a été estimé à partir de la moyenne de trois ratios similaires à l'incidence obtenus à partir des données relatives aux donneurs anglais entre 2005 et 2007 (nouveaux donneurs et donneurs connus) [21]. Il est important de noter que, dans la simulation, l'incidence n'a pas été calculée selon sa définition formelle. C'est plutôt la probabilité qu'un don provienne d'un donneur incident au VIH ($P_{incident}$) qui a été utilisée. On obtient une probabilité qu'une valeur en personnes par année. On a

$$P_{incident} = \frac{\text{nombre de cas incidents}}{\text{nombre de dons testés}}.$$

Afin de calculer la prévalence, le nombre de cas de dons provenant de donneurs prévalents est divisé par le nombre total de dons testé pour chacun des sexes dans chacune des catégories de donneurs (nouveaux ou connus) sur une période donnée soit la période de 5 ans mentionnée plus haut. Pour l'incidence, le nombre de cas incidents est divisé par le nombre total de dons testés du 1^{er} juillet 2014 au 30 juin 2019.

Le taux de dons incidents et le taux de prévalence du VIH chez les HARSAH ont d'abord été estimés pour les donneurs connus en fonction de la compliance. L'estimation de la

prévalence et de la proportion de dons incidents chez les donneurs HARSAH est plus complexe puisque l'on a peu de données. Il est donc nécessaire de faire certaines suppositions. Pour le calcul de la prévalence, il y a 3 cas prévalents dont un incident. On prend alors pour acquis que le cas incident provient d'un donneur non compliant et, en outre, qu'il s'agit du seul cas de non compliance. Toutefois, le donneur aurait pu être infecté plus de trois mois avant son don et être quand même incident. Si ce donneur n'a pas donné au cours des 3 dernier mois, alors ce serait un donneur compliant. Puisqu'il y a un seul don provenant d'un HARSAH incident, on ne tient pas compte de la compliance au critère HARSAH dans le calcul de l'incidence du VIH chez ces donneurs. Aussi, on considère qu'un donneur équivaut à une personne/année. Puisque le taux moyen de dons par an par donneur est de 2,15, on multiplie le nombre de donneur HARSAH par cette valeur afin d'obtenir un résultat en fonction du nombre de dons HARSAH par an. Le taux moyen de dons par année faits par un donneur à été estimé par les organismes responsables de la collecte des produits sanguins au Canada.

On estime que les taux de dons incidents et de prévalence du VIH chez les nouveaux donneurs HARSAH sont dix fois plus élevés que les taux estimés pour les donneurs connus [12]. Les valeurs de taux d'incidence et de prévalence utilisés dans la simulation sont présentés dans le tableau 3.3.

Une fois que le statut infectieux du don est déterminé, une charge virale lui est associée en fonction du type d'infection, que ce soit une infection de type prévalent ou incident ou encore une infection de type incident en période silencieuse.

Tableau 3.3 : Probabilité de dons incidents et prévalence du VIH

Prévalence du VIH		
Sexe	Statut du donneur	Valeur
Homme	Nouveau donneur	0.724 par 100,000 dons
	Donneur connu	0.117 par 100,000 dons
Femme	Nouveau donneur	0.613 par 100,000 dons
	Donneur connu	0.306 par 100,000 dons
Taux d'incidence du VIH		
Sexe	Valeur	
Homme	0.039 par 100,000 dons	
Femme	0.102 par 100,000 dons	
Prévalence du VIH chez les donneurs HARSAH		
Statut du donneur	Compliance	Valeur
Nouveau donneur	Non-compliant (ou HARSAH à risque élevé)	155.40 par 100,000 dons
	compliant	65.28 par 100,000 dons
Donneur connu	Non-compliant (ou HARSAH à risque élevé)	15.54 par 100,000 dons
	compliant	6.53 par 100,000 dons
Taux d'incidence du VIH chez les donneurs HARSAH		
Statut du donneur	Compliance	Valeur
Donneur connu	Non-compliant (ou HARSAH à risque élevé)	15.54 par 100,000 dons
	compliant	2.7 par 100,000 dons

3.2.3 Charge virale

Afin d'estimer la charge virale, deux méthodes ont été utilisées puisqu'il faut diviser la distribution de la charge virale en fonction du temps en deux périodes principales. Soit la période avant et la période suivant la séroconversion. La séroconversion est le moment où il devient possible de détecter la présence du virus dans le sang. On nomme période silencieuse, le délai entre l'infection et la séroconversion. Ce délai varie selon le test de dépistage utilisé. Dans la présente recherche, le TAN est utilisé et la période silencieuse (PS) est de $9(\pm 0,6)$ jours. La charge virale durant la PS est indétectable, c'est-à-dire qu'elle est trop faible pour être mesurée. On a donc aucune donnée pour quantifier la charge virale au cours de cette période. Le modèle présenté par Weusten et al. [2] à été utilisé afin d'estimer la virémie au cours de cet intervalle de temps. On considère que la charge virale augmente de manière exponentielle au cours de la période silencieuse suivant un temps de doublement λ . La concentration C de virus dans le sang en virions/mL se calcule avec l'équation suivante :

$$C = C_0 2^{\frac{t}{\lambda}},$$

où C_0 est la concentration initiale lors du moment d'infection, λ est le temps de doublement et t est le temps écoulé depuis le moment d'infection.

Les valeurs utilisées dans les calculs présentés dans la simulation se retrouvent dans le tableau 3.4.

Tableau 3.4 : Données associées au calcul de la virémie durant la période silencieuse

Variable	Valeur	Distribution	Référence
λ	0,85 (0,76-0,97) jours	Uniforme	Weusten et al.[2]
C_0	0,00229093 virion	Estimation ponctuelle	Weusten et al.[2]
t	[0, longueur PS]	uniforme	

Afin de déterminer le temps écoulé depuis le moment d'infection t , une distribution uniforme est utilisée : $\mathcal{U}(0, L_{PS})$ où L_{PS} est la longueur de la période silencieuse en jours.

De manière à simuler la probabilité qu'un don infecté ait été fait durant la période silencieuse, l'intervalle interdon médian est considéré. Il s'agit du temps médian entre deux dons de plasma chez un même donneur. La durée de la période silencieuse est l'autre variable utilisée dans le calcul. Cette dernière varie selon une distribution uniforme $\mathcal{U}(8.4, 9.6)$. La probabilité qu'un don infecté soit fait durant la période silencieuse est donc :

$$P(\text{Don fait durant la PS}) = \frac{\text{Durée de la PS}}{\text{Longueur de l'intervalle interdon}}.$$

L'estimation de la charge virale pour les dons infectés faits en dehors de la PS, est effectuée à partir de la distribution de la charge virale de dons infectés estimée à partir d'une base de données fournie par le SRAP. Il s'agit de données provenant de dons sanguins faits aux États-Unis et ayant testé positif au VIH. C'est à partir de ces données que la distribution de la charge virale post période silencieuse a été modélisée. À première vue, en regardant l'histogramme, on pourrait croire que la distribution des données suit une courbe log-normale.

Cependant, on remarque un pic de valeurs près de zéro. Afin de tester la normalité des données log-transformées, une technique de validation graphique a d'abord été utilisée : le diagramme quantile-quantile. L'image qui suit (figure 3.4) présente le résultat associé aux données de charge virale.

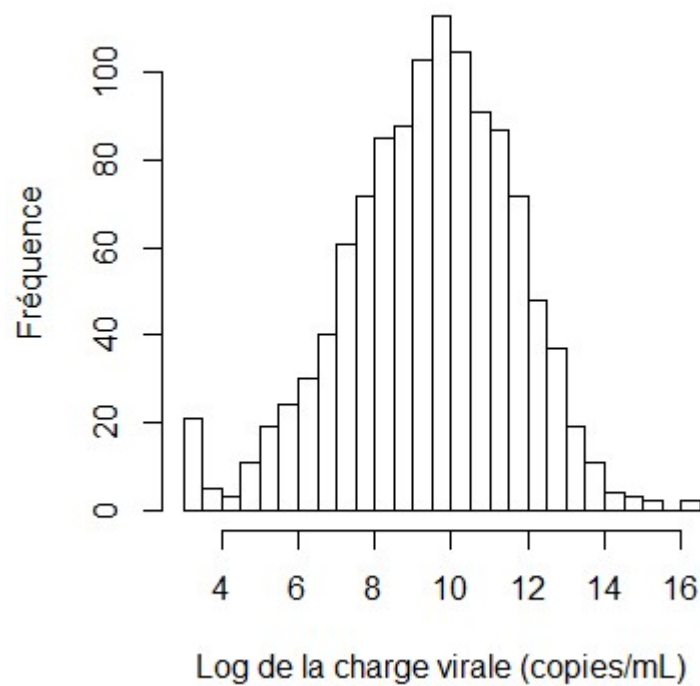


FIGURE 3.3 – Histogramme de la charge virale.

Sur la figure 3.4, la droite représente la distribution de référence et le nuage de points, les données log-transformées. On remarque qu'aux extrémités, les données s'éloignent de la droite. Cela semble indiquer qu'il n'y a pas d'adéquation entre les données transformées et la distribution gaussienne. Afin de confirmer l'hypothèse précédente, le test de shapiro-Wilk est utilisé pour les données auxquelles une transformation logarithmique a été appliquée.

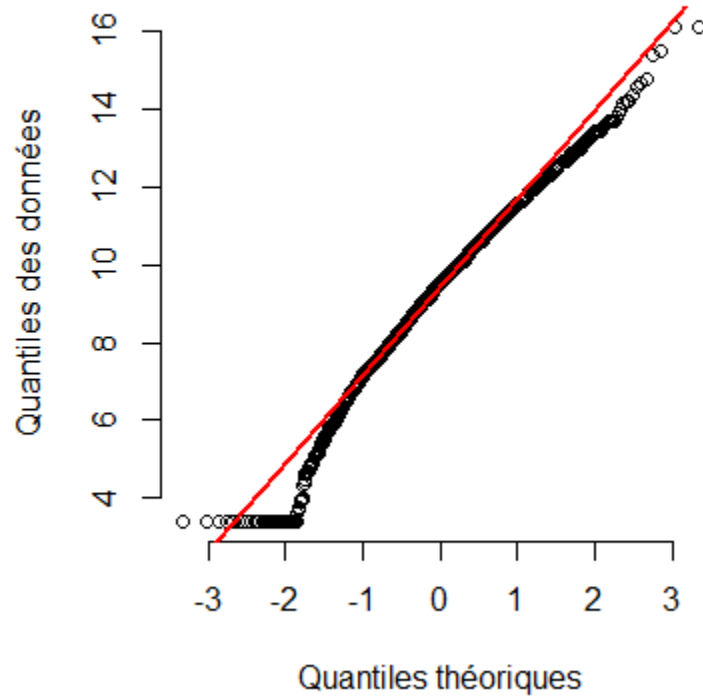


FIGURE 3.4 – Diagramme quantile-quantile de la charge virale.

Une statistique W est obtenue avec une p-valeur de 6.648×10^{-11} , ce qui est très près de zéro et donc on rejette l'hypothèse de la normalité. Puisque l'on cherche une estimation fidèle aux données, on ne peut associer la charge virale à aucune distribution paramétrique. Une approche d'estimation de densité non-paramétrique est plutôt utilisée. Tout d'abord, la densité de l'ensemble des données est estimée avec l'approche de l'estimateur à noyau. La courbe de densité présentée sur la figure 3.5 est obtenue. L'écart entre la plus petite et la plus grande valeur de charge virale est très grand. Effectivement, la plus petite donnée charge virale est de 30 copies/ml alors que la plus élevée est de 10000000 copies/ml. De plus, la base de donnée contient très peu de valeurs très élevées. Il devient

donc difficile d'estimer adéquatement la densité des données.

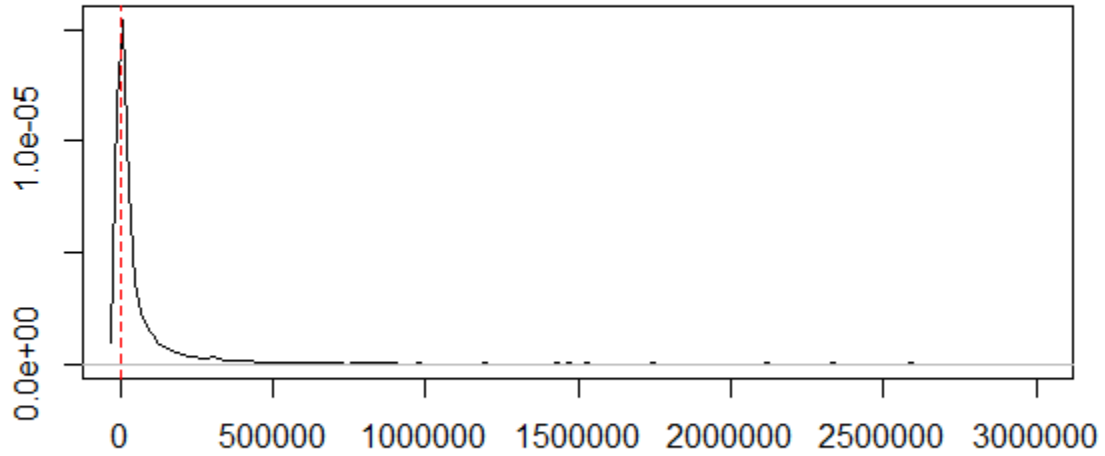


FIGURE 3.5 – Densité de la charge virale basée sur les données non catégorisées.

Il est possible de remarquer, sur la figure 3.5, que la densité admet un poids positif sur des valeurs négatives, ce qui est impossible pour une charge virale puisqu'il s'agit d'une concentration qui ne peut être négative. De plus, les valeurs extrêmes que l'on retrouve dans la queue de la distribution pourraient être sous-estimées.

Les données ont donc été divisées en cinq intervalles basés sur la charge virale en copies/mL, soit les cinq catégories suivantes : $[0, 10000)$, $[10000, 50000)$, $[50000, 150000)$, $[150000, 500000)$ et $[500000, \infty)$. La densité liée aux valeurs de charge virale dans chacune des catégories a été estimée avec la méthode d'estimateur à noyau. Toutefois, pour

les deux intervalles extrêmes, soit $[0, 10000)$ et $[500000, \infty)$, la transformation logarithmique présentée au chapitre précédent a été nécessaire afin de représenter adéquatement les données et de respecter les bornes, le cas échéant. Le noyau gaussien a été utilisé pour l'estimation de la densité dans chacun des intervalles. Pour les trois intervalles centraux, le calcul du paramètre de lissage s'est fait à l'aide de la règle de Silverman. Pour les deux intervalles dont les données ont subi une transformation, la méthode de calcul du paramètre de lissage introduite par Sleather et Jones (1991) a été utilisée. Dans tous les cas, plusieurs techniques d'estimation du paramètre de lissage (validation croisée, règle de Silverman, méthode de Sleather et Jones) ont été comparées afin de choisir le paramètre optimal. La proportion de valeurs dans chacune de ces catégories a été considérée afin que l'estimation reflète bien les données. La figure suivante présente la distribution associée à la charge virale en copies/ml dans chacune des catégories.

Une fois que l'on a estimé la distribution de la charge virale, il est possible d'associer une valeur de CV à un don infecté que l'on a préalablement simulé. Si le don infecté a été effectué durant la période silencieuse, la méthode de *Weusten et al.*[2] est utilisée. Sinon, un intervalle parmi les cinq est déterminé aléatoirement en fonction de la probabilité de se retrouver dans chacune des catégories. Une valeur de charge virale estimée est associée au don infecté en fonction de la distribution de la virémie de l'intervalle. L'étape suivante consiste à simuler l'effectivité du TAN. Une fois que suffisamment de dons sont simulés afin de constituer un lot, la technique de réduction des pathogènes est modélisée.

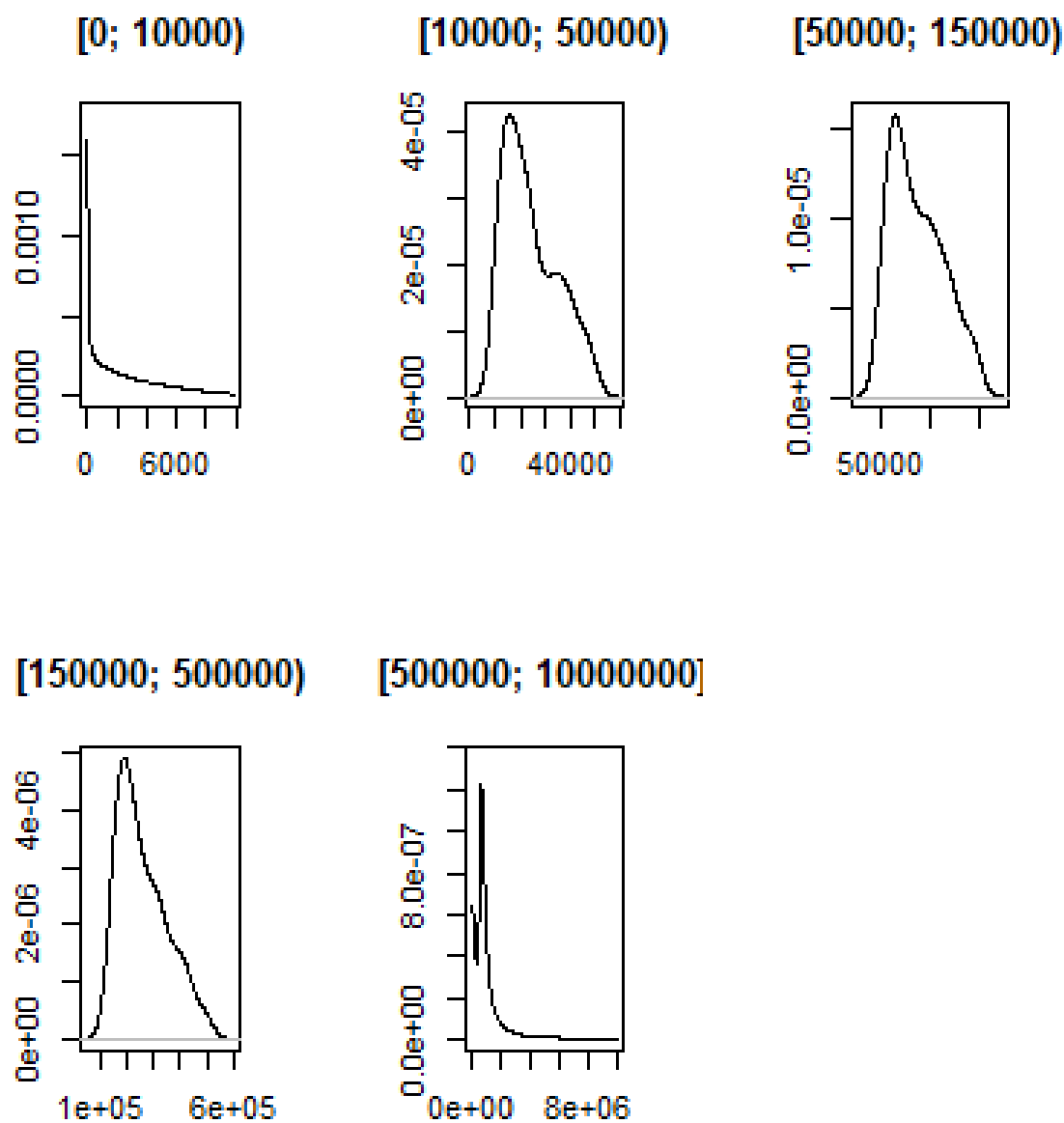


FIGURE 3.6 – Densité de la charge virale par catégories.

3.2.4 Test d'acide nucléique et technologie de réduction des pathogènes

Depuis janvier 2001, les tests d'acides nucléiques sont utilisés par Héma-Québec et la SCS [30, 31]. Il est connu que, lors de la contraction d'un virus, ce dernier n'est pas immédiatement identifiable par des tests, c'est la période silencieuse du virus. Le TAN permet de réduire significativement la durée de cette phase. En effet, avec les tests sérologiques qui vérifient la présence d'antigènes et d'anticorps associés au VIH, la période silencieuse se trouvait à quelque part entre trois semaines et trois mois. Les tests de quatrième génération détectent l'antigène p24 et les anticorps liés au VIH. On estime que l'infection est détectée à 50% 18 jours après l'exposition au virus et qu'elle est détectée à 99% après 44 jours. Depuis la mise en place du TAN, la période silencieuse est plutôt d'une durée variant de sept à quatorze jours [32].

La Société canadienne du sang et Héma-Québec utilisent deux TAN différents. Chez la SCS, on utilise le test Cobas Multiplex (ROCHE MOLECULAR SYSTEMS) en bassin de six unités. Chez Héma-Québec, on utilise le Procleix Ultrio Plus (Genprobe Inc) en bassin de huit unités. Chaque test a une sensibilité, une spécificité et un taux d'échec qui lui sont propres. Puisque la présente étude considère l'ensemble des dons de plasma faits au Canada, sans distinction liée l'organisme qui a effectué la collecte, la valeur de sensibilité ainsi que la probabilité d'échec du test sont obtenues dans l'encart de chacun des tests et les valeurs les plus élevées sont utilisées dans les simulations. On retrouve les différentes valeurs dans le tableau 3.5. Deux autres types d'erreurs ont été considérés, soit l'erreur de transcription et l'erreur clinique. L'erreur de transcription est associée aux produits sanguins qui ont testé positif, mais qui se retrouvent dans l'inventaire. L'erreur clinique est le taux de faux négatifs associés à une erreur de manipulation [22].

Une distribution a priori a été associée à chacune de ces probabilités d'erreur afin de modéliser les intervalles de confiance et d'inclure de l'incertitude sur les valeurs. Les quatres erreurs susmentionnées ont été combinées afin de déterminer la probabilité d'échec du TAN. Lors de la simulation, on modélise le TAN une fois qu'un don est simulé et que son statut infectieux est déterminé. La figure 3.7 présente la partie du GOA tiré du réseaux bayésien de la simulation associé à cette étape de la simulation.

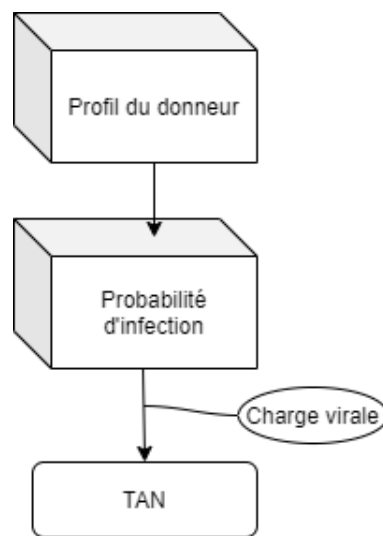


FIGURE 3.7 – GOA associé au TAN.

Tous les dons identifiés comme ayant été prélevés durant la période silencieuse d'un individu infecté sont automatiquement considérés comme non détectés par le TAN. Les dons infectés au VIH et détectés par le TAN ne se retrouvent pas dans le bassin de dons de plasma destiné au fractionnement. Par contre, tous les dons infectés que le TAN a échoué à détecter s'y retrouveront. La charge virale associée aux dons infectés se retrouvant dans le bassin de plasma est diluée dans l'ensemble des dons de plasma pour ensuite être réduite par le processus de réduction des pathogènes appliqué au plasma de fractionnement. Ce processus est assez complexe et nécessite plusieurs étapes. La log

réduction est utilisée comme mesure pour quantifier la diminution de concentration virale.

La log réduction (R) est évaluée à partir des concentrations initiale (C_0) et finale (C) d'un contaminant. On l'évalue avec la formule suivante :

$$R = -\log_{10} \left(\frac{C}{C_0} \right).$$

Alors,

$$10^{-R} = \frac{C}{C_0}.$$

On en déduit le pourcentage de réduction P associé à R :

$$P = 100 \times \frac{(C_0 - C)}{C_0} = 100 \times (1 - 10^{-R}).$$

Le tableau suivant présente les pourcentages de réduction associés à différentes valeurs de R .

Tableau 3.5 : Log réductions

Valeur de R	Réduction (%)
1	90
2	99
3	99,9
4	99,99
5	99,999
6	99,9999

Plus la valeur de R est élevée, plus le pourcentage de réduction est grand.

Plusieurs procédés appliqués au plasma destiné au fractionnement mènent à une réduction des pathogènes. Il y a la pasteurisation, le traitement par solvant/détergent, le traitement à la chaleur sèche ainsi que le traitement à la vapeur chaude des produits lyophilisés. Chacune des techniques précédentes cible soit les virus enveloppés comme le VIH, soit les virus non-enveloppés comme le virus de l'hépatite A (VHA), ou encore les deux à la fois.

Dans le cadre de la simulation, la valeur de $R = 9,6$ utilisée est basée sur les log réductions associées à différents produits thérapeutiques issus du plasma (PTP) présentées dans un communiqué de CSL Behring, une compagnie biopharmaceutique spécialisée dans les PTP. Les valeurs de log réduction varient entre 9,6 et 15,3 pour le VIH parmi tous les produits fabriqués par la compagnie CSL Behring. La valeur la plus faible (9,6) a été utilisée. Une log réduction de 9,6 équivaut à un pourcentage de réduction de 99,9999997488%.

Finalement, le tableau 3.6 présente les valeurs liées au TAN et à la TRP.

Quelques scénarios présentant une variation de certains paramètres comme les erreurs liées au TAN, à la proportion de HARSAH dans la population de donneurs, etc., ont été simulés afin de comparer différents résultats avec le scénario de base puisque les paramètres sont principalement obtenus à partir d'estimations. Ces scénarios sont présentés en détail dans la section suivante.

Tableau 3.6 : Variables associées à la TRP et au TAN

Paramètre	Valeur	Référence
Période muette (jours) ²	9 (±0,6)	O'Brien et al. [12]
Intervalle interdon median (jours)	14	Basé sur les données de HQ et de la SCS
Sensibilité du TAN	100% (94,0-100,0) ³	Notice d'accompagnement Cobas MPX [13]
Taux d'échec du TAN ¹	0,00% (0,00-1,22) ³	Notice d'accompagnement Cobas MPX [13]
Erreur de transcription ¹	1:10 000 000 (1:4 000 000–1:100 000 000) ³	Germain et al. [5]
Erreur clinique ¹	1:100 000 (1: 35 500–1: 1 000 000) ³	Germain et al. [5]
PRT log-reduction	9,6	Déclaration de CSL Behring

¹Variabilité induite avec une distribution beta

²Variabilité induite avec une distribution normale

³ IC 95%

3.2.5 Différents scénarios

Puisque les données liées aux HARSAH sont difficiles à récolter dans certaines situations, la simulation est basée principalement sur des valeurs estimées pour, la proportion de donneurs HARSAH, la prévalence et l'incidence du VIH chez les HARSAH, la compliance

au critère d'exclusion ou encore la proportion d'HARSAH à risque élevé dans le cas où il n'y aurait aucune exclusion. Il y a donc beaucoup d'incertitude associée aux données. Pour cette raison, quatre scénarios faisant varier les valeurs des paramètres mentionnés ont été considérés pour chacun des modèles d'exclusion HARSAH. Il y a le scénario le plus probable, le scénario optimiste et deux scénarios pessimistes. Un scénario extrême a été ajouté pour le modèle sans exclusion seulement.

1. **Scénario le plus probable** : Les valeurs d'incidence et de prévalence du VIH chez les HARSAH, la compliance ou encore les individus HARSAH à haut risque ainsi que la proportion de donneurs HARSAH utilisées sont celles que l'on retrouve dans les tableaux 3.7 et 3.8.
2. **Scénario optimiste** : Les valeurs d'incidence, de prévalence et la proportion de donneurs HARSAH demeurent les mêmes que dans le scénario le plus probable. Le taux de non-compliance chez les hommes pour le modèle considérant une exclusion HARSAH de trois mois ainsi que la proportion de donneurs HARSAH à risque élevé dans le cas d'un modèle sans exclusion ont été diminué de moitié.
3. **Scénarios pessimistes** : Pour les deux scénarios pessimistes, l'incidence et la prévalence du VIH chez les HARSAH, le taux de non-compliance ou la proportion de HARSAH à risque élevé sont doublés par rapport au scénario le plus probable. La différence entre les deux cas pessimistes est au niveau de la proportion de donneurs HARSAH.
 - (a) Dans le premier scénario pessimiste, on multiplie par un facteur de 1.5 le nombre de donneurs HARSAH, ce qui augmente la proportion de donneurs HARSAH comparativement au scénario le plus probable.
 - (b) Pour le deuxième scénario pessimiste, on considère plutôt l'estimation de la proportion d'HARSAH dans la population générale d'hommes au Canada [33] afin de transposer cette valeur à la proportion de donneurs HARSAH.

Les valeurs associées aux paramètres HARSAH variant en fonction des scénarios sont présentées dans le tableau 3.7.

Suite à la suggestion d'experts et afin d'estimer le pire risque possible, un scénario extrême a été modélisé. Dans ce scénario, la proportion d'hommes HARSAH dans la population générale est utilisée comme estimation du nombre de donneurs HARSAH. Les taux d'incidence et de prévalence du VIH dans la population HARSAH générale sont aussi utilisées comme valeurs d'incidence et de prévalence chez les donneurs HARSAH [34]. Ces valeurs sont nettement plus élevées que celles retrouvées dans la population de donneurs HARSAH actuelle.

Les résultats des simulations associés à chacun des scénarios se retrouvent dans la prochaine section.

Tableau 3.7 : Valeurs des différents paramètres HARSAH selon les scénarios

Paramètres			Scénarios			
			Le plus probable	Optimiste	Pessimiste	
					(i)	(ii)
Proportion de donneurs HARSAH	Modèle d'exclusion HARSAH de 3 mois	0,016274 (0,009992- 0,025424)	0,016274 (0,009992-0,025424)	0,03254 (0,019984- 0,050848)	0,029 (0,018- 0,039)	
	Modèle sans exclusion HARSAH	0,017257 (0,010981- 0,026398)	0,017257 (0,010981-0,026398)	0,034514 (0,021962- 0,052796)	0,029 (0,018- 0,039)	
Incidence du VIH chez les HARSAH	Nouveau donneur	15,54	15,54	31,08		
	Donneur connu	2,7	2,7	5,4		
Prévalence du VIH chez les HARSAH	Nouveau donneur	Non- compliant	155,4	155,4	310,8	
		Compliant	65,28	65,28	130,56	
	Donneur connu	Non- compliant	15,54	15,54	31,08	
		Compliant	6,53	6,53	13,06	
Taux de non-compliance chez les hommes pour le modèle d'exclusion HARSAH de 3 mois		0,001 (0,0003-0,002)	0,0005 (0,00015-0,001)	0,002 (0,0006-0,004)		
Proportion d'hommes à risque HARSAH élevé pour le modèle sans exclusion HARSAH		0,001998 (0,0005998- 0,00399)	0,000999 (0,0002999- 0,001995)	0,003996 (0,0011996-0,00798)		

Tableau 3.8 : Valeurs des paramètres HARSAH pour le scénario extrême [33, 34]

Proportion de HARSAH	Prévalence du VIH	Incidence du VIH
0.029 [0.018-0.039]	0.0960487	0.0095245

3.2.6 Résultats

Ici, les risques associés au scénario le plus probable pour chacun des modèles d'exclusion sont d'abord comparés. Ensuite, les différences entre le deuxième scénario pessimiste pour chacun des modèles d'exclusion sont dépeintes. Finalement, les résultats obtenus pour le scénario extrême sont discutés. Il est intéressant de noter que, pour chacun des scénarios, 300 000 lots contenant chacun entre 4 000l et 6 000l de plasma ont été simulés. Un don de plasma récolé par aphérèse a un volume de 750ml. Chaque lot de plasma contient donc entre 5 333 et 8 000 dons. Il est nécessaire de faire une comparaison entre les résultats obtenus suite aux simulations sous les deux hypothèses de politiques HARSAH afin d'estimer le risque lié à un tel changement de politique. Les résultats suivants font l'objet d'un article accepté pour la publication *Vox Sanguinis*. Ce dernier est présenté à l'annexe A.

Pour le scénario le plus probable du modèle d'exclusion de trois mois, qui représente le risque actuel, la nombre moyen de copies du VIH dans chacun des lots de plasma à la suite du processus de réduction des pathogènes est de $9,7606 \times 10^{-8}$ (IC 95% [$9,7598 \times 10^{-8}$; $9,7615 \times 10^{-8}$]) copies/lot, la médiane est de 1.7536×10^{-8} et la concentration maximale est de 1.7672×10^{-6} copies/lot. Comparons ces résultats avec ceux obtenus lors de la simulation du scénario le plus probable du modèle sans exclusion HARSAH : on a obtenu une moyenne de $9,7856 \times 10^{-8}$ (IC 95% [$9,7847 \times 10^{-8}$; $9,7865 \times 10^{-8}$]) copies par lot, une médiane de 1.4700×10^{-8} copies/lot ainsi qu'une valeur maximale de 2.2000×10^{-6} copies/lot. On constate qu'il y a une variation de 0,256% pour la moyenne. De plus, la probabilité d'obtenir un lot infecté sous scénario le plus probable du modèle 3-mois est de 0,00441 comparativement à 0,00494 pour le modèle sans exclusion. Cela équivaut à une augmentation de 12,02%. Le nombre moyen de copies du virus par milliard de litres passe de 0,019334 copie à 0,019511 d'un modèle à l'autre ce qui équivaut à une différence

de 0,000177 copie par milliard de litres.

Lors de la simulation du deuxième scénario pessimiste (ii), c'est-à-dire celui où l'on considère que la proportion de donneurs masculins HARSAH serait égale à la proportion de HARSAH dans la population générale, les résultats obtenus sont les suivants : des simulations basées sur le modèle dans lequel on considère l'hypothèse de non exclusion des HARSAH, la probabilité d'obtenir un lot infecté par le VIH est de 0,01331. Un nombre moyen de $1,0007 \times 10^{-7}$ (IC 95% [$1,0006 \times 10^{-7}$; $1,0008 \times 10^{-7}$]) copies/lot par lot à été obtenu. Les concentrations médiane et maximale sont respectivement de $1,6400 \times 10^{-8}$ copies/lot et de $2,2800 \times 10^{-6}$ copies/lot. La simulation du scénario pessimiste (ii) associé au modèle d'exclusion HARSAH de trois mois a donné les résultats suivants : une concentration virale moyenne de $9,7277 \times 10^{-8}$ (IC 95% [$9,7269 \times 10^{-8}$; $9,7285 \times 10^{-8}$]) copies/lot, un nombre de copie par lot médian de $1,7200 \times 10^{-8}$ et maximal de $1,9900 \times 10^{-6}$ copies/lot. Les variations absolues entre les deux modèles sont respectivement de 2,87%, 4,65% et 14,57% pour les nombre de copies par lot moyens, médians et maximaux. Par milliard de litres, on note une différence de 0,000412 copies entre les concentrations moyennes.

Examinons maintenant les résultats obtenus lors de la simulation du scénario extrême. Sous les hypothèses considérées dans ce scénario, les simulations démontrent que la probabilité d'obtenir un lot contaminé au VIH est de 0,02872. Le nombre moyen de copies par lot est de $1,6295 \times 10^{-6}$ (IC 95% [$1,6244 \times 10^{-6}$; $1,6346 \times 10^{-6}$]). Le nombre médian est de $1,6300 \times 10^{-8}$ copies/lot et le nombre maximal est de 0,0104 copies/lot. Les valeurs sont légèrement plus élevées que dans les autres scénarios en partie parce que trois dons infectés par le VIH et faits à l'extérieur de la période silencieuse se sont retrouvés dans des lots.

Le risque d'obtenir un lot de plasma de fractionnement positif au VIH suite à l'appli-

cation d'un processus de réduction des pathogènes est minime. En effet, si on considère qu'un lot est infectieux s'il contient au moins une copie du VIH, aucune des simulations de 300 000 lots de plasma, ce qui représente environ 2 milliards de dons, n'a donné lieu à un lot infectieux. En effet, le risque d'obtenir un lot infectieux est très faible, soit de 0,00 avec un intervalle de confiance à 95% de $[0,00; 1,229618924 \times 10^{-5}]$.

Les résultats associés à chacun des scénarios pour les deux modèles sont présentés dans le tableau 3.9.

Tableau 3.9 : Résultats selon les scénarios

	Modèle d'exclusion	PLUS PROBABLE	OPTIMISTE	PESSIMISTE 1	PESSIMISTE 2	CATASTROPHE
Dons VIH positifs par 1,000,000 dons	3 mois aucune	2.79 3.01	2.71 2.86	5.10 5.96	5.68 6.32	--- 57.4
Nombre de lots infectés (sur 300,000 lots)	3 mois aucune	1323 1483	1259 1335	2885 3719	3136 3994	--- 8617
Concentration virale moyenne après TAN et TRP (copies/ml)	3 mois aucune	1.93×10^{-14} 1.95×10^{-14}	1.82×10^{-14} 1.66×10^{-14}	1.95×10^{-13} 1.95×10^{-14}	1.97×10^{-14} 2.01×10^{-14}	--- 2.81×10^{-13}
Probabilité qu'un lot soit infecté	3 mois aucune	0.00441 0.00494	0.00420 0.00445	0.00962 0.01240	0.01045 0.01331	--- 0.02872
Nombre moyen de copies par lot après TAN et TRP (copies/lot)	3 mois aucune	9.76×10^{-8} 9.79×10^{-8}	9.17×10^{-8} 8.16×10^{-8}	9.75×10^{-8} 9.70×10^{-8}	9.73×10^{-8} 1.00×10^{-7}	--- 1.63×10^{-8}
Nombre médian de copies par lot après TAN et TRP (copies/lot)	3 mois aucune	1.75×10^{-8} 1.47×10^{-8}	1.55×10^{-8} 1.49×10^{-8}	1.65×10^{-8} 1.59×10^{-8}	1.72×10^{-8} 1.64×10^{-8}	--- 1.63×10^{-8}
Nombre maximal de copies par lot après TAN et TRP (copies/lot)	3 mois aucune	1.77×10^{-8} 2.20×10^{-8}	1.15×10^{-8} 1.26×10^{-8}	2.05×10^{-8} 2.00×10^{-8}	1.99×10^{-8} 2.28×10^{-8}	--- 0.01041

3.3 Analyse de sensibilité

Dans les chapitres précédents, le processus de modélisation qui mène à l'estimation du risque résiduel d'obtenir un lot de plasma de fractionnement contaminé est présenté en détail. La difficulté d'obtenir certaines données nécessaires à l'estimation est considérée. Certains des paramètres ont donc dû être estimés et une certaine variabilité leur a été accordée à l'aide de distributions a priori. Plusieurs scénarios ont été simulés impliquant la variation de plusieurs paramètres simultanément. Il est toutefois nécessaire de vérifier la robustesse du modèle en effectuant une analyse de sensibilité permettant d'identifier les paramètres les plus susceptibles d'influencer l'estimation du risque. Ce chapitre présente l'analyse de sensibilité dans son entièreté.

Afin d'effectuer une analyse de sensibilité, il est d'abord nécessaire d'identifier les paramètres les plus susceptibles d'avoir un impact sur l'estimation du risque ou encore les paramètres dont l'estimation de la valeur pourrait être variable en fonction de certains facteurs. Ces paramètres ainsi que leur variation ont été déterminés en collaboration avec les membres du SRAP.

Afin de mesurer l'impact d'une erreur d'estimation des paramètres, chacun de ces paramètres a été soumis, individuellement, à certaines variations. De nouvelles estimations du risque ont été obtenues pour chacune de ces variations. Les paramètres visés par cette analyse étaient les valeurs de prévalence du VIH chez les donneurs HARSAH connus ainsi que chez les nouveaux donneurs HARSAH et les nouveaux donneurs HARSAH non-compliants au critère d'exclusion, les taux d'incidence du VIH dans la population de donneurs HARSAH connus et dans la population de donneurs HARSAH connus non-compliants au critère d'exclusion, le taux d'échec du test d'acide nucléique ainsi que

le taux de faux négatifs de ce dernier. Les variations appliquées à chacun de ces paramètres sont présentées dans le tableau suivant. L'analyse de sensibilité a été effectuée en comparaison avec le scénario le plus probable associé au modèle d'exclusion de trois mois.

L'analyse de sensibilité a permis d'identifier certains facteurs ayant un impact plus marqué sur l'estimation du risque résiduel dans le présent contexte.

Tableau 3.10 : Analyse de sensibilité : Variations appliquées aux paramètres

Paramètres	Valeur initiale	Valeur modifiée
Prévalence VIH nouveaux donneurs MSM non-compliants (par 100,000 dons)	155.40	210.80
Prévalence VIH nouveaux donneurs MSM (par 100,000 dons)	65.28	130.56
Prévalence VIH donneurs MSM connus non-compliants (par 100,000 dons)	15.54	31.08
Prévalence VIH donneurs MSM connus (par 100,000 dons)	6.53	13.06
Incidence VIH donneurs MSM connus non-compliants (par 100,000 dons)	15.54	31.08
Incidence VIH MSM connus (par 100,000 dons)	2.7	5.4
Sensibilité du TAN (% de faux négatifs)	0.00 [0.00 ; 0.018]	0.00 [0.00 ; 0.036]
		0.00 [0.00 ; 0.072]
Taux d'échec du TAN (%)	0.00 [0.00 ; 1.22]	0.00 [0.00 ; 2.44]
		0.00 [0.00 ; 4.88]

D'abord, les résultats de l'analyse démontrent que le fait de doubler la valeur de l'incidence chez les donneurs HARSAH connus, soit de passer d'une probabilité de 0.000027 à une probabilité de 0.000054 d'obtenir un don incident au VIH, a eu l'effet d'augmenter

le nombre de dons positifs au VIH effectués durant la période silencieuse et donc mené à une augmentation du nombre de lots de plasma présentant une charge virale non nulle.

Ensuite, l'analyse de sensibilité a permis de conclure que le fait de doubler ou de quadrupler le taux d'échec du TAN ou encore le taux de faux négatifs liés à la sensibilité du test n'a eu aucun impact sur l'estimation du risque. Une modification de la valeur de prévalence au VIH chez les donneurs HARSAH connus n'a pas un impact significatif sur l'estimation du risque d'obtenir un lot de plasma contaminé. En effet, doubler le taux de prévalence chez les donneurs HARSAH connus augmente le nombre de dons prévalents effectués, mais ces derniers ne se retrouvent pas dans un lot de plasma dû à la robustesse des tests d'acide nucléique. La modification de paramètres associés aux HARSAH non-compliants n'a pas non plus d'effet manifeste sur le risque. Ce résultat est probablement lié au taux de non-compliance particulièrement faible. En effet, puisque la proportion des donneurs HARSAH non-compliants est très faible par rapport à la population de donneurs générale, une augmentation du taux de donneurs infectés dans cette catégorie n'a pas d'influence marquée sur le risque.

Aucune des modifications précédentes n'a eu l'effet de produire un lot infectieux. Toutefois, il est important de noter que, sans l'utilisation de la technologie de réduction des pathogènes, une augmentation de la concentration de la virémie dans un lot pourrait être problématique.

Les résultats de l'analyse de sensibilité sont présentés dans le tableau 3.11.

Tableau 3.11 : Résultats de l'analyse de sensibilité

	Scénario le plus probable (3 mois)	Prévalence VIH chez les nouveaux donneurs HARSAH non-compliants x_2^1	Prévalence VIH chez les nouveaux donneurs HARSAH x_2^2	Prévalence VIH chez les donneurs HARSAH connus non-compliants x_2^3	Prévalence VIH chez les donneurs HARSAH connus x_2^4	Incidence VIH chez les donneurs HARSAH connus non-compliants x_2^1	Incidence VIH chez les donneurs HARSAH connus x_2^1	Sensibilité du TAN		Probabilité d'échec du TAN	
								$(x_2)^1$	$(x_4)^1$	$(x_2)^1$	$(x_4)^1$
Dons positifs au VIH pour 1,000,000 dons	2.79	2.82	2.91	2.87	3.52	2.79	2.72	2.79	2.79	2.79	2.79
Nombre de lots de plasma présentant une charge virale (sur 300,000 lots)	1323	1302	1314	1236	1260	1323	1696	1323	1323	1323	1323
Probabilité d'obtenir un lot infecté	0.00441	0.00434	0.00438	0.00412	0.00420	0.00441	0.00565	0.00441	0.00441	0.00441	0.00441
Nombre de copies moyen par lot après TAN et à la TRP ($\times 10^8$ copies)	9.7606	10.138	26.636	181.49	9.3127	9.7606	9.2143	9.7606	9.7606	9.7606	9.7606
Nombre de copies médian par lot après TAN et à la TRP ($\times 10^8$ copies)	1.7536	1.8600	1.7450	1.5800	1.6350	1.7536	1.7350	1.7536	1.7536	1.7536	1.7536
Nombre de copies maximal par lot après TAN et à la TRP ($\times 10^6$ copies)	1.7672	1.6600	147.39	213.23	2.0300	1.7672	1.7000	1.7672	1.7672	1.7672	1.7672

¹ En comparaison avec le scénario le plus probable du modèle d'exclusion HARSAH de 3 mois.

CONCLUSION

L'étude présentée dans ce mémoire a pour but de quantifier le risque résiduel de contamination au VIH d'un lot de plasma traité par un processus d'inactivation des pathogènes destiné au fractionnement et composé de dons de plasma récoltés par aphérèse au Canada dans l'éventualité de l'abolition du critère d'exclusion du don de trois mois depuis la dernière relation sexuelle pour les HARSAH. Cette évaluation du risque est, entre autres, une réponse à la demande des défenseurs de droits de la communauté LGBTQ2 visant à abolir le traitement différentiel associé aux HARSAH en lien avec le don de sang. L'estimation du risque est justifiée par les taux de prévalence et d'incidence du VIH plus élevées chez les HARSAH et afin d'assurer la sécurité des receveurs.

Avant d'arriver à un résultat, les différents paramètres pouvant influencer le risque ont dû être déterminés. Des dépendances ont été identifiées entre certains paramètres. Ces dernières ont été prises en compte dans la modélisation. Il a été nécessaire de faire certaines estimations à partir de bases de données ou encore à partir de données retrouvées dans la littérature afin d'attribuer des valeurs à chacun des paramètres du modèle. Une fois toutes les données récoltées, des simulations de type Monte Carlo à partir d'une structure de réseau bayésien ont été effectuées à l'aide du logiciel statistique R.

Tel que discuté dans la section résultats du chapitre précédent, aucune des simulations n'a donné lieu à un lot infectieux. Les résultats indiquent que le risque de transmission du VIH dans les produits thérapeutiques dérivés du plasma est négligeable.

Les tests d'acide nucléique sont très robustes, ainsi la probabilité de ne pas détecter un don fait à l'extérieur de la période silencieuse est presque nul. Les résultats des simulations démontrent bien ce fait. En effet, pour le scénario catastrophe, on retrouve seulement trois dons effectués en dehors de la période silencieuse se retrouvant dans des lots pour 2 000 004 185 dons et aucune de ces erreurs n'est attribuée à un échec du TAN. Cela équivaut à trois erreurs de procédure par 10 000 ans au Canada.

Dans l'éventualité d'une modification à la politique d'exclusion HARSAH, l'évaluation de l'admissibilité au don serait plutôt basée sur une étude des comportements à risque réalisée à l'aide d'un questionnaire auquel seraient soumis les donneurs potentiels. On y retrouverait des questions telles que : *Au cours des trois derniers mois, avez-vous eu une relation sexuelle avec un homme ? ; Avez-vous, au cours des 12 derniers mois, habité ou eu une relation sexuelle avec une personne atteinte de l'hépatite ? ; Avez-vous, au cours des 12 derniers mois, eu une relation sexuelle avec un individu ayant déjà consommé des drogues illégales par injection ? ; Avez-vous, au cours des 12 derniers mois, eu une relation sexuelle avec un individu ayant reçu de l'argent ou de la drogue en échange de services sexuels ? ; Avez-vous, au cours des 12 derniers mois, reçu de l'argent ou de la drogue en échange de services sexuels ? ; Avez-vous eu la gonorrhée au cours des 12 derniers mois ? ; Avez-vous, au cours des 12 derniers mois, eu une relation sexuelle avec un individu ayant le SIDA ou ayant testé positif au VIH ?.*

Dans le futur, il serait intéressant de faire une étude du risque basée sur les différents

comportements à risque des donneurs HARSAH. Il pourrait aussi être possible d'étendre la simulation à différents produits sanguins, tels les plaquettes, ainsi qu'à d'autres virus tels le VHB et le VHC en simulant l'efficacité de la technologie de réduction des pathogènes INTERCEPT® qui est approuvée par santé Canada entre autres pour les plaquettes.

Finalement, il serait sécuritaire de permettre aux HARSAH sexuellement actifs de donner du plasma. De plus, un tel changement apporterait une hausse de 1,05% des donneurs de plasma, ce qui est intéressant puisque la demande pour les produits thérapeutiques dérivés du plasma est en hausse au Canada [35].

ANNEXE A

Cette Annexe présente l'article associé au troisième chapitre de ce mémoire qui a été accepté par la revue Vox Sanguinis. Cet article a été co-écrit avec plusieurs scientifiques internationaux membres du SRAP. Leurs noms et affiliations sont présentés sous le titre.

HIV Residual Risk In Canada For Apheresis Source Plasma Donation Without Deferral For Men Who Have Sex With Men

Eliana Aubé^{1,2}, Antoine Lewin^{2,3}, Sheila F. O'Brien⁴, Yves Grégoire⁵, Josiane Pillonel⁶, Whitney R. Steele⁷, Brian Custer^{8,9}, Katy L. Davison¹⁰, Marc Germain⁵, Clive R. Seed¹¹, Felix Camirand Lemyre^{1,12}

On behalf of the Surveillance, Risk Assessment and Policy Subgroup of the ISBT Transfusion Transmitted Infectious Diseases Working Party

Author affiliation

1. Department of Mathematics, Université de Sherbrooke, Sherbrooke, QC, Canada
2. Medical Affairs and Innovation, Héma-Québec, Montreal, QC, Canada

3. Faculty of Medicine and Health Science, Université de Sherbrooke, Sherbrooke, QC, Canada
4. Epidemiology & Surveillance, Canadian Blood Services, Ottawa, ON, Canada
5. Medical Affairs and Innovation, Héma-Québec, Quebec City, QC, Canada
6. Direction des Maladies Infectieuses, Santé publique France, St. Maurice, France
7. Scientific Affairs, American Red Cross, Rockville, MD, USA
8. Epidemiology and Health Policy Science, Vitalant Research Institute, San Francisco, CA, USA
9. Department of Laboratory Medicine, UCSF, San Francisco, CA, USA
10. Immunisation, Hepatitis & Blood Safety, Health Protection Agency, London, UK
11. Donor and Product Safety (DAPS) Policy Unit, Australian Red Cross Lifeblood, Perth, WA, Australia
12. Centre de recherche, Centre hospitalier universitaire de Sherbrooke, Sherbrooke, QC, Canada

Correspondence & Reprints : Félix Camirand Lemyre Faculté des sciences, Université de Sherbrooke 2500 boulevard de l'Université Sherbrooke, Qc (Canada) J1K 2R1 E-mail : felix.camirand.lemyre@usherbrooke.ca

Financial support : This research received funding support from Canadian Blood Services MSM Research Grant program, funded by the federal government (Health Canada) and the provincial and territorial ministries of health, from the National Sciences and Engineering Research Council of Canada and the Fond de recherche du Québec – Nature et Technologies. Australian governments fund Australian Red Cross Lifeblood to provide blood, blood products and services to the Australian community.

Running title : HIV RESIDUAL RISK AND APHERESIS SOURCE PLASMA

Key words : HIV, Residual Risk, MSM, Deferral, Pathogen Inactivation Technology, Canada

Acknowledgements : All authors contributed to the design, reviewed model outputs and contributed to interpretation. EA performed the analysis and drafted the manuscript and all authors revised it critically. All authors approved of the submitted version.

Conflict of interest : The authors declare that there are no conflicts of interest identified.

Word count : 4990.

ABSTRACT

BACKGROUND AND OBJECTIVES : In Canada, men having sex with men (MSM) are deferred for 3 months from last sexual contact to reduce human immunodeficiency virus (HIV) risk to recipients. The aim of this paper was to model the Canadian residual risk of HIV-positive source plasma incorporating pathogen inactivation (PI) under no MSM deferral scenarios for apheresis plasma donations.

MATERIALS AND METHODS : A combined Bayesian network and Monte Carlo approach was implemented to estimate the HIV residual risk under three-month deferral compared with no deferral without quarantine scenarios for MSM donors. Models involve the stochastic generation of donation and its infection status based on its corresponding simulated donor profile. Viral load reduction conferred by PI used by source plasma fractionators was simulated. Model parameters were derived from Héma-Québec and Canadian Blood Services data, viral loads in a large sample of HIV-positive US blood donors, CSL Behring documentation, and from published data.

RESULTS : In the most likely scenario for the three-month deferral model, there were 2.71 positive donations per 1,000,000 donations (95% CI 2.63 – 2.78). For the no-deferral model, there were 3.01 positive donations per 1,000,000 donations (95% CI 2.94 – 3.09). For both scenarios, the risk of having an infectious pool was 0 in 300,000 pools (95% CI 0 – 0.0000123) after consideration of PI.

CONCLUSION : Based on simulation results, there would be a negligible HIV residual risk associated with the removal of a time-based MSM deferral without quarantine for source plasma incorporating PI.

INTRODUCTION

Since human immunodeficiency virus (HIV) emerged in the early 1980s, major efforts have been made to ensure blood products' safety and quality in order to avoid transmission of infectious agents to recipients [1]. One approach has been to exclude from donation those at higher risk for HIV. In North America, beginning in the early 1980s, men having sex with men (MSM) was identified as a very-high-risk group for acquired immunodeficiency syndrome (AIDS). Although little was known about HIV as the causative agent of AIDS at the time, MSM were permanently deferred from blood donation if they had sex with another man even once, since 1977. Since then, scientific knowledge concerning HIV infection, diagnostics and transmission modes have considerably expanded. Sensitive assays have reduced the window period (WP) of infection and computer systems have all but eliminated the chance of releasing a product in error. Internationally, MSM deferral from blood donation was initially permanent, but numerous countries have changed the deferral from a lifetime to progressively shorter periods [2]. Based on development of highly accurate testing methods and introduction of pathogen inactivation (PI), completely lifting the ban on donation from sexually active MSM has been suggested [3]. In particular, it has been proposed that MSM could safely donate plasma intended for fractionation, given that each donation is screened by serology and nucleic acid testing (NAT), and that the fractionation process includes very effective PI steps. However, it is necessary to carefully assess risks associated with this change since this group of potential donors remains at higher risk of HIV infection in Western countries. For example, at the end of 2016, the Public Health Agency of Canada estimated that 49.1% of HIV-infected persons in Canada were MSM [4].

In Canada, the lifetime exclusion for MSM blood donors was abolished in 2013 and replaced by a 5-year deferral. In 2016, deferral changed from 5 years to 12 months, and, in

2019, Canada authorized MSM blood donations after 3 months since last sexual contact with another man [5]. These policy changes were supported by mathematical models simulating various scenarios which allowed estimating the risk of releasing potentially infectious donations [5, 6]. Based on these estimations, simulated HIV residual risk remains extremely low, given that sensitivity of NAT has improved to the extent that the HIV WP for donation testing in Canada is 8–10 days [5].

One approach now being considered in Canada is the possible implementation of less restrictive MSM deferral for plasma intended for fractionation. In this context, we developed a stochastic BN model with Monte Carlo simulations to estimate the risk of HIV contamination of plasma-derived products if sexually active MSM were allowed to donate.

MATERIALS AND METHODS

Context

We define HIV residual risk as the probability of having an infectious pool of plasma after using PI processes. The infectious threshold is conservatively assumed to be one HIV RNA copy per pool. A pool designates the combination of a number of individual plasma donations required to reach a predefined volume of plasma. Individual donations are assumed to originate from donors who are considered safe to donate based on their answers to numerous risk questions, including those regarding sexual risks (see Data S1) and whose donation has tested negative for HIV-1/2 antibody and HIV-1 NAT. This allows an estimate of what the residual risk would be in a model where the MSM deferral for source plasma is 3 months (current) versus no MSM deferral (proposed).

Model

A two-stage Monte Carlo procedure was employed to estimate the rate of HIV-positive

pools in N=300,000 pools of fractionated plasma treated with one or more PI steps.

First stage : A BN is a probabilistic graphical model that uses a directed acyclic graph (DAG) to specify conditional relationships among a collection of variables. In the BN DAG, nodes represent variables, and edges reflect conditional dependencies. A joint probability distribution can then be constructed from the graphical structure by elicitation of the set of conditional probabilities of each node given its parents in the graph [7]. Figure 1 shows the BN DAG considered for our simulation models. It includes donor variables known to be associated with risk of infection : age at the time of donation, sex, donor status (new/repeat), MSM (yes/no), compliance with MSM deferral (yes/no, in the 3-month scenario only), level of risk (high/low, in the no-deferral scenario only), HIV infectious state (not infected/infected within the WP/infected and not in WP) and NAT (+/-). According to the BN graph in Figure 1, the joint probability over these variables factors as :

$$\begin{aligned}
P(age, sex, MSM, status, compliance, HIV state, NAT) = \\
P(age)P(sex | age)P(MSM | sex) \times \\
P(status | age, sex, MSM)P(compliance | MSM) \times \\
P(HIV state | sex, MSM, status, compliance)P(NAT | HIV state).
\end{aligned}$$

The Elicitation section below contains a description of how bodies of knowledge were used to model each conditional probability factor above.

Second stage : The Monte Carlo algorithm performed simulations of each donation individually and independently from the joint distribution specified by the BN. To form plasma pools, simulated plasma donations (750 ml) were iteratively combined until a pooled volume ranging randomly from 4000 L to 6000 L was reached (personal communication with Rima Khalil, Héma-Quebec Stable products director). The HIV viral load

of a pool was calculated as the sum of the viral load of each positive donation adjusted for the volume of each plasma unit and the total volume of the pool. The effect of the PI process was simulated based on the viral load of each pool.

Elicitation

Simulation process flowchart

The BN DAG in Figure 1 was used as a process flowchart for donor profile simulations. First, age was generated, second, sex conditional on age, third, MSM status conditional on sex and so on. In the no-deferral model, rate of non-compliance was used as a proxy for the percentage of recently active high-risk MSM potential donors (calculations are shown in the Scenarios section). Although not explicitly shown in Figure 1, infectious state was also set to depend on prevalence and incidence rates. These were made specific to sex, donor status, MSM status and compliance variables. For newly positive donations, an extra simulation step was carried to consider WP risk.

Parameters Used in the First Stage

Donor age and sex. Distribution of donors by age group ($P(\text{age})$) was set to match sample proportions calculated based on plasma donation data from Canadian Blood Services (CBS) and Héma-Québec collected in 2018 (January 1st to December 31st). The minimum age for donation is 17 at CBS and 18 at Héma-Québec. Three age groups were considered : 17/18 to 24, 25 to 44 and over 44. This stratification was made since it could be seen from the data that the proportion of new donors (which needed to be simulated in our model) varied significantly among these three groups. Distribution of sex within each age group ($P(\text{sex} | \text{age})$) was determined similarly.

MSM donor proportion. $P(\text{MSM} = \text{yes} | \text{sex} = \text{female})$ was set to 0 and $P(\text{MSM} |$

sex = male) was adjusted based on the MSM male donor proportion for the 12-month deferral model described in O'Brien et al., 2019 [8], see Data S1.

Donor status. The distribution of new and repeat donors ($P(status | age, sex, MSM)$) was elicited by matching sample proportions from Héma-Québec and CBS internal data.

MSM donor non-compliance with MSM deferral criteria. The rate of MSM failing to comply with deferral criteria was determined based on male non-compliance rate retrieved from the most likely scenario of O'Brien et al. 2020 [5]. Non-compliance in MSM ($P(compliance = no | MSM = yes)$) was estimated using the non-compliance rate in men and MSM proportion : $P(compliance = no | MSM = yes) = (P(non-compliance in men)) / (P(MSM = yes))$ Further, $P(compliance = no | MSM = no)$ was set to 0.

HIV state. For the conditional distribution of HIV state, parameters associated to HIV incidence/prevalence and WP probability were considered.

HIV prevalence/incidence : Here, prevalence is defined as the proportion of positive donations in a donation pool, whereas the incidence rate is defined as the proportion of new HIV cases in a donation pool over a year. Given the limited number of HIV- positive donations in our datasets (three positive donations in five years), prevalence and incidence rate estimations were not categorized by age group and were determined for all groups of donors $G = G_{(sex, MSM, status, compliance)}$ corresponding to a plausible choice for donor profile variables.

To determine group-wise HIV prevalence and incidence rates, donations in Canada (CBS and Héma-Québec) from July 1, 2014, to June 30, 2019 were considered. For groups matching status=repeat, prevalence was elicited according to the formula

$$\text{Prevalence in } G_{\text{status=repeat}} = \frac{\text{number of HIV infected donations in } G_{\text{status=repeat}}}{\text{number of donations tested in } G_{\text{status=repeat}}}.$$

Prevalence in groups with status=new was considered 10 times higher than the one of their matched analogue with *status = repeat* (i.e. same characteristics except for status) obtained from the above equation [5].

To determine group-wise HIV incidence, an annual mean of 2.15 donations per donor was used to estimate the number of annual donations. Group-specific incident cases rate was determined from HIV-positive donations made by repeat donors who had made an HIV-negative donation in the 12 months preceding their positive donation. For groups associated with status=repeat, incidence rate was calculated as :

$$\text{Incidence rate in } G_{\text{status=repeat}} = \frac{\text{number of incident cases in } G_{(\text{status} = \text{repeat})}}{\text{number of donations tested in } G_{\text{status=repeat}}}.$$

Incidence rate in groups with status=new was elicited following the approach of Davison et al. 2019, and set equal to the one of their matched analogue with status=repeat multiplied by a truncated-at-zero normal random variable with mean of 1.65 (SD 1.0334). This adjustment factor was put forward following the analysis of first time and repeat English donor data during 2001-2007 [6].

For cases where *MSM = yes*, prevalence and incidence proportions were set equal to the number of HIV- positive donations from non-compliant MSM donors (current criteria) and the estimated number of non-compliant MSM donors was based on post donation interviews and compliance survey.

WP donations : WP is the stage from infection to time at which pathogen becomes detectable by a given diagnostic procedure. The probability of having an incident donation being made during the WP was considered independent of the donor’s profile given their “newly infected” status, and set equal to :

$$P(\text{WP donation} \mid \text{incident case} = \text{yes}) = \frac{\text{WP length}}{\text{Median interdonation interval length}}.$$

The median interdonation interval length was calculated based on 2018 Canadian data for plasma donation (CBS and Héma-Québec).

Parameters used in the second stage

Viral load. For WP donations, Weusten et al. model [9] was used to specify the viral load distribution. The maximum WP viral load was set to 14.5 copies/mL. Viral loads post WP were simulated based on U.S. HIV-positive blood donation database (1,156 positive blood donations from the period of 2010 to 2018). Complete modelling details regarding viral load distribution are given in Data S1.

HIV serological and nucleic acid testing. Performance, limit of detection and system failure parameters were elicited based on routinely used procedures in Canadian blood operators involving Cobas multiplex and Procleix Ultrio Plus package inserts [10, 11], see Data S1 for details.

PI process. Donor screening and testing of donations and fractionation pools are the first and second steps in product safety, while virus reduction using PI is the third and last step applied to pooled plasma destined for plasma-derived medical products [12] (see also Data S1). The log reduction factor (LRF) is used to quantify this reduction, for instance, one LRF corresponds to inactivating 90% of the pathogen with the viral load being

reduced by a factor of 10 (i.e., 1 log). HIV LRF for PDMD ranges from 15.3 (Privigen, CSL Behring, Bern, Switzerland) to 9.6 (RiaSTAP, CSL Behring, Marburg, Germany) (Personal communication with CSL Behring). For modelling purposes, the lowest mean LRF for HIV was used.

Model parameter values related to NAT performance and PI log reduction are detailed in Data S1.

Scenarios

Simulations were conducted according to a three-month MSM deferral model and to a model with no MSM deferral without plasma unit quarantine. Both were simulated under four scenarios : most likely, optimistic, and pessimistic i and ii. Values for the parameters are shown in Tables 1-3.

1. **Most likely scenario** : HIV incidence, prevalence, MSM donor proportion and non-compliance rates were taken as described in the Elicitation section. For the no-deferral model, proportion of males with high-risk MSM behaviour was determined using the proportion of non-compliant male donors as proxy for high-risk MSM behaviour as follows : $\frac{(non-compliantmaledonorproportion(3-month) \times 2)}{(1 + non-compliantmaledonorproportion(3-month))}$.

The three-month non-compliant male donor proportion was multiplied by 2 for the no-deferral model since MSM donors who were non-compliant before lifting of the deferral were considered to remain blood donors after the lifting, and the number of high-risk newly eligible MSM donors was considered to be proportional to that number.

2. **Optimistic scenario** : HIV incidence, prevalence and MSM donor proportion were

set as in the most likely scenario. Non-compliance in male donors for three-month deferral and high-risk MSM proportion for no-deferral model were reduced by half.

3. Pessimistic scenarios :

- i Compared to the most likely scenario, incidence, prevalence, non-compliance and/or high-risk MSM proportion were doubled, and MSM donor proportion was increased by a factor of 1.5.
- ii Parameters were set as in pessimistic-i, except for MSM donor proportion. For the latter, the estimation of MSM proportion within the general population in Canada of 2.9% (95% CI 1.8-3.9) was used, combined with the donor rate calculated for the general population [13].

In addition, a worst-case scenario was simulated for the no-deferral model using MSM donor proportion and MSM HIV incidence and prevalence rates based on the MSM general population (i.e. 443 per 100,000 MSM persons and 9.60%, respectively) [14, 15].

RESULTS

In the most likely scenario for the three-month deferral model, the probability of having a pool with HIV RNA copy was 0.00441. An average number of HIV copies per pool after use of PI of 9.7606×10^{-8} copy (95%CI $9.7598 \times 10^{-8} \sim 9.7615 \times 10^{-8}$) was obtained, with a median of 1.7536×10^{-8} copy and a maximum of 1.7672×10^{-6} copy. For the no-deferral scenario, probability of having a pool with HIV RNA copy was 0.00494. The number of HIV copies per pool was 9.7856×10^{-8} copy (95%CI $9.7847 \times 10^{-8} \sim 9.786510^{-8}$), along with a median number of 1.4700×10^{-8} and a maximum number of 2.2000×10^{-6} copy per pool. The mean number of HIV copy per billion liters ranges from 0.019334 copy to 0.019511 corresponding to a difference of 0.000177 copy per billion liters between the

averages of the two deferral models.

For the worst-case scenario, the probability of having a pool with HIV RNA copy was 0.02872. The average number of HIV copies per pool after use of PI was of 1.6294×10^{-6} copy (95%CI $1.6244 \times 10^{-6} \sim 1.6346 \times 10^{-6}$), with a median of 1.6300×10^{-8} copy and a maximum of 0.010409 copy. Results are presented in Table 4. There, absolute differences between mean viral concentrations are smaller than 1.59×10^{-15} copy/mL and are thus not clinically significant. Finally, under the one-HIV RNA copy-per-pool assumed threshold, none of the 300,000 pools simulation rounds, each round representing around two billion donations, would have been considered infectious (i.e. no manufactured product would have contained one virion of HIV after the use of PI process). A sensitivity analysis was conducted to highlight variables that are likely to affect risk estimates (Data S2).

DISCUSSION

Our model allowed to simulate HIV contamination risk in pools of plasma with a change from 3 months to no deferral without quarantine for MSM, considering the use of PI steps in the fractionation process. The results suggest that the latter risk of HIV transmission in plasma-derived products is negligible. According to our model, allowing sexually active MSM to donate would not significantly increase the risk of occurrence of an HIV RNA-positive pool. Our results also suggest that, while the vast majority of pools would not contain HIV at all, the average viral load within the HIV RNA-positive pools would only be slightly higher under the no-deferral model compared to the one under the 3-month deferral model and would be well under the one-per-pool HIV RNA copy threshold. Therefore, the safety profile of plasma-manufactured products, with respect to HIV transmission, would likely remain unaffected. Of note, our assumed threshold is arguably conservative, see, for example, Zanetti et al. [16], which states, regarding plasma

for transfusion, that one human infectious dose corresponds to approximately 500–1000 HIV RNA copies. See also the study of HIV acquisition in Kleinman et al. [17].

In the worst-case scenario, only three infectious donations made outside the WP ended up in a pool, none of these, due to an NAT failure. This corresponds to three procedural errors in every 10,000 years in Canada.

Each source of data employed to estimate the model parameters has its limitations, and some parameters were more challenging to estimate. For instance, a response bias might have been present in surveys used to estimate the non-compliance rate among men for the MSM deferral criteria [18]. Non-compliance in Canada is low and may be different in other countries [19]. Our estimation of these rates was based on the ones presented by O’Brien et al. [8], obtained from a Canadian survey, to reflect Canadian donor behaviour with greater accuracy. Another limitation is that this simulation study did not consider the pre-exposure prophylaxis (PrEP) medication used to reduce the burden of HIV. PrEP use could interfere with HIV screening among blood donors and may result in undetectable RNA by nucleic acid/viral load assays for blood component qualification [20]. PrEP medication use is a deferral criterion for blood donors in Canada, and hence, its inclusion in the model should arguably not significantly increase the proportion of donations with HIV viral loads below the NAT detection limit under a realistic compliance rate assumption.

Despite the need to generate evidence-based data supporting policy changes regarding MSM ensuring the safety of recipients, few studies have evaluated intention to donate plasma among MSM. A qualitative study performed by Grace et al. indicated that time-based deferral would maintain the key issues regarding equity and fairness related to blood screening practices for MSM in Canada [21]. According to participants, policies regarding MSM blood donations should be more aligned with sound and up-to-date scientific evidence. Our results strongly suggest that allowing sexually active MSM to

donate plasma for fractionation would be just as safe as the current approach of time-limited deferral, without having to impose a potentially inefficient and costly quarantine strategy that would be perceived as unduly burdensome and discriminatory.

Demand for plasma-derived medicinal products in Canada has increased over the past 10 years [22]. Our study supports the safety of expanding source plasma donation to include donors not subject to deferral for MSM. Based on the number of source plasma donations in Canada in 2018 (106,654 donations), no deferral model would lead to an increase of about 1119 donations (1.05%) if we added newly eligible donors according to pessimistic ii scenario.

Based on our modelling, there would be a negligible HIV contamination risk associated with a change from 3-month MSM deferral to no MSM deferral for source plasma donors in Canada. Given the aim of implementing less restrictive MSM deferral policies in Canada, our results support that it would be a safe next step to permit MSM to donate source plasma without a time-based deferral.

REFERENCES

1. Perkins HA, Busch MP. Transfusion-associated infections : 50 years of relentless challenges and remarkable progress. *Transfusion*. 2010 ;50 :2080-2099.
2. Goldman M, Shih AW-Y, O'Brien SF, et al. Donor deferral policies for men who have sex with men : past, present and future. *Vox Sanguinis*. 2018 ;113 :95-103.
3. Villa C H. Pathogen Reduction of Platelet Donations as an Alternative Procedure to MSM Donor Deferral Blood Products. Advisory Committee Meeting March 21, 2019. [Accessed Jan 7, 2021].
URL : <https://www.fda.gov/media/127506/download>
4. Public Health Agency of Canada. Summary : Estimates of HIV incidence, prevalence and Canada's progression on meeting the 90-90-90 HIV targets, 2016. 2018.

5. O'Brien SF, Grégoire Y, Pillonel J, et al. HIV residual risk in Canada under a three-month deferral for men who have sex with men. *Vox Sang.* 2020 ;115 :133-139.
6. Davison KL, Gregoire Y, Germain M, et al. Changing the deferral for men who have sex with men – an improved model to estimate HIV residual risk. *Vox Sang.* 2019 ;114 :666-674.
7. Koski T, Noble JM. *Bayesian Networks : An Introduction.* Wiley ; 2009.
8. O'Brien SF, Osmond L, Fan W, et al. Compliance with time-based deferrals for men who have sex with men. *Transfusion.* 2019 ;59 :916-920.
9. Weusten J, Vermeulen M, van Drimmelen H, et al. Refinement of a viral transmission risk model for blood donations in seroconversion window phase screened by nucleic acid testing in different pool sizes and repeat test algorithms. *Transfusion.* 2011 ;51 :203-215.
10. Cobas MPX test package insert.[Accessed Dec 16, 2020]. URL :
[https ://www.fda.gov/media/115031/download](https://www.fda.gov/media/115031/download)
11. Procleix Ultrio Plus assay package insert. [Accessed Dec 16, 2020]. URL :
[https ://www.fda.gov/media/73892/download](https://www.fda.gov/media/73892/download)
12. Klamroth R, Gröner A, Simon TL. Pathogen inactivation and removal methods for plasma-derived clotting factor concentrates : Plasma Concentrate Pathogen Safety. *Transfusion.* 2014 ;54 :1406-1417.
13. Pillonel J, Pelat C, Tiberghien P, et al. The evolving blood donor deferral policy for men who have sex with men : impact on the risk of HIV transmission by transfusion in France. *Transfusion.* 2020 ;60 :525-534.
14. Rich AJ, Lachowsky NJ, Sereda P, et al. Estimating the Size of the MSM Population in Metro Vancouver, Canada, Using Multiple Methods and Diverse Data Sources. *J Urban Health.* 2018 ;95 :188-195.

15. Challacombe L. The epidemiology of HIV in gay men and other men who have sex with men. Catie Fact Sheet. 2015. [Accessed 2020 Nov 29].
URL : <https://www.catie.ca/ga-pdf.php?file=sites/default/files/epi-msm%20EN%202015%2002%2025b.pdf>
16. Zanetti AR, Bodini U, Clerici M, et al. Transfusion of red blood cells from an HIV-RNA-positive/anti-HIV-negative donor without HIV infection in the recipient. *Transfusion*. 2007 ;47 :1328-9.
17. Kleinman SH, Lelie N, Busch MP. Infectivity of human immunodeficiency virus-1, hepatitis C virus, and hepatitis B virus and risk of transmission by transfusion. *Transfusion*. 2009 ;49 :1328-9.
18. Seed CR, Lucky TT, Waller D, et al. Compliance with the current 12-month deferral for male-to-male sex in Australia. *Vox Sang*. 2014 ;106 :14-22.
19. Sauvage C, Spinardi R, Pelat C, et al. Noncompliance with blood donor selection criteria - Complidon 2017, France. *Transfusion*. 2020 ;60 :73-83.
20. Custer B, Quiner C, Haaland R, et al. HIV antiretroviral therapy and prevention use in US blood donors : a new blood safety concern. *Blood*. 2020 ;136 :1351-8.
21. Grace D, Gaspar M, Lessard D, et al. Gay and bisexual men's views on reforming blood donation policy in Canada : a qualitative study. *BMC Public Health*. 2019 ;19 :772.
22. Hotchko M, Robert P. Recent market status and trends of fractionated plasma products. *Ann Blood*. 2018 ;3 :19-19.

TABLES AND FIGURES

Table 1 – MSM parameter values according to deferral model

		Three-month deferral (95% CI)	No deferral (95% CI)
Proportion of MSM among donor population		0.0162 (0.0099 – 0.0254)	0.0172 (0.0110 – 0.0264)
Proportion of non-compliant MSM		0.0010 (0.0003 – 0.0020)	-----
Proportion of men at high risk of HIV		-----	0.0020 (0.0006 – 0.0040)
Proportion of MSM among new donors by age group	17/18-24	0.0522 (0.0348 – 0.0843)	0.0562 (0.0398 – 0.0864)
	25-44	0.0279 (0.0225 – 0.0324)	0.0332 (0.0282 – 0.0375)
	>44	0.0158 (0.0135 – 0.0186)	0.0218 (0.0197 – 0.0245)

Table 2 – HIV prevalence and incidence according to deferral models

HIV prevalence in the donor population		
Sex	Donor type	HIV prevalence per 100,000 donations
Men	First time	0.724
	Repeat	0.117
Women	First time	0.613
	Repeat	0.306
Probability that donation is an incident case for HIV in donor population		
Sex	HIV incident donation rate per 100,000 donations	
Men	0.039	
Women	0.102	
MSM HIV prevalence		
Donor type	Compliance	HIV prevalence per 100,000 donations
First time	Non-compliant or high risk	155.40
	MSM donor population	65.28
Repeat	Non-compliant or high risk	15.54
	MSM donor population	6.53
MSM HIV incident donation probability		
Donor type	Compliance	HIV incident rate per 100,000 donations
Repeat	Non-compliant or high risk	15.54
	MSM donor population	2.7

Table 3 – Parameter values for WP and interdonation interval

Parameter	Value	Reference
Window period (mean (SD)) ¹	9 (0.6) days	O'Brien et al. [5]
Interdonation interval (median [IQR; Q1, Q3])	14 days [7 – 34]	HQ and CBS

¹Variability modelled by a normal distribution.

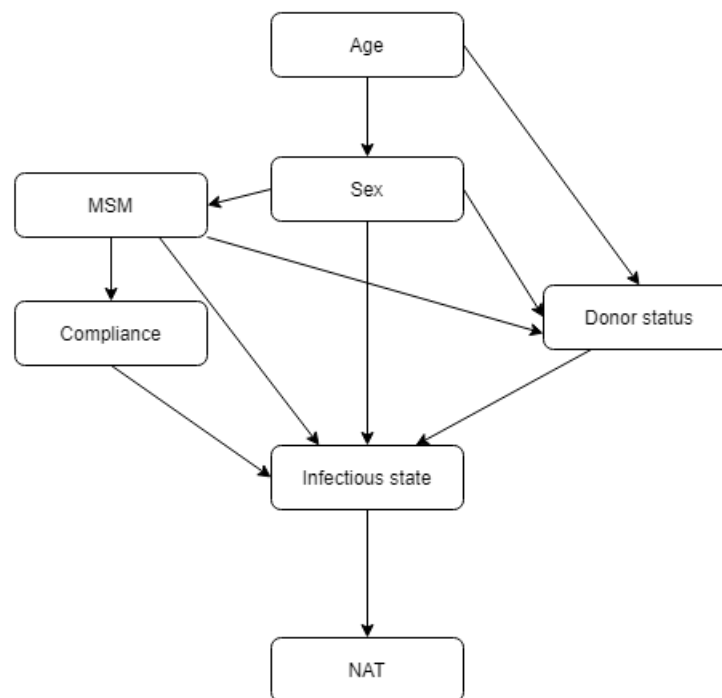


Figure 1 – DAG of HIV risk estimation for source plasma

Data S1

Risk Questions, Men/Women/MSM Proportion Probabilities, Prior Elicitation, Viral Load Density Estimation and Additional details regarding HIV serological/nucleic acid testing and PI steps

Risk questions : In the manuscript, it is stated that “[i]ndividual donations are assumed to originate from donors who are considered safe to donate based on their answers to numerous risk questions”. The questions considered establishing donors risk status include, for women : in the last three months, have you had sex with a man who, in the last 12 months, had sex with a man ?. For men, the questions include : have you had sex with a man ? ; In the last 3 months, have you had sex with a man who, in the last 12 months, had sex with a man ? ; In the last 3 months, have you had sex with a man ? ; In the last 12 months, have you resided or had sex with a person who has had hepatitis ? ; In the last 12 months, have you had sex with someone who has ever taken illegal drugs by injection ? ; In the last 12 months, have you had sex with someone who has taken money or drugs for sex ? ; In the last 12 months, have you had gonorrhea ? ; In the last 12 months, have you had sex with someone who has AIDS or has tested positive for HIV ? ; In the last 12 months, have you taken money or drugs for sex ?

Men, Women and MSM proportion probabilities : The proportion of additional MSM donors for 3-month deferral model ($P_{addMSM\ 3M}$) was approximated using the proportion of MSM who had sex with another man in the six months prior to donation ($P_{MSM\ sex\ 6M}$), in the past year excluding last six months ($P_{MSM\ sex\ year-6M}$) and proportion of MSM who had sex in the last 3 months ($P_{MSM\ sex\ 3M}$) :

$$P_{addMSM\ 3M} = P_{MSM\ sex\ 6M} + P_{MSM\ sex\ year-6M} - P_{MSM\ sex\ 3M}$$

MSM donor proportion under a 3-month deferral ($P_{MSM\ donor\ 3M}$) was based on MSM donor proportion under a 12-month deferral ($P_{MSM\ donors\ 12M}$) :

$$P_{MSM\ donor\ 3M} = \frac{P_{MSM\ donors\ 12M} + P_{add\ MSM\ donors\ 3M}}{1 + P_{add\ MSM\ donors\ 3M}}$$

The MSM donor proportion in no- deferral model ($P_{MSM\ donor\ 0M}$) was set to :

$$P_{MSM\ donor\ 0M} = \frac{P_{MSM\ donor\ 3M} + P_{MSM\ sex\ 3M}}{1 + P_{MSM\ sex\ 3M}}$$

Modelling of uncertainty : Prior probability distributions were assigned to each proportion parameter to better reflect degree of uncertainty in their elicitation. Based on the donation data collected by the two Canadian blood operators, overall and monthly average proportions were estimated for donors according to age and sex. The overall average was used as the mean proportion and minimum and maximum values for monthly average proportions were used, respectively, as lower and upper bounds of a 95% confidence interval. A beta distribution was associated to each proportion, and its parameters (α, β) were chosen based on the quartiles of the associated confidence interval. The beta distribution was used as prior probability distribution for the proportions mostly because it is defined on the $[0,1]$ interval like a proportion and its parameters can be chosen to reflect data. Appendix S1 Table 1 presents age group, men by age group and women by age group proportions and their associated 95% confidence interval. Prior distribution parameters were chosen according to these quartiles. **Viral load probability density estimation** : Viral loads for the 1,156 HIV positive specimens were obtained using the Hologic (San Diego, CA) Aptima HIV-1 quantitative viral load assay which has a dynamic range of 30 copies/mL to 10,000,000 copies/mL [1]. Median viral load concentration for positive donations outside WP was 14,684.5 copies/mL [IQR 55,597.5 copies/mL]. Viral load probability density was estimated using kernel estimation techniques, since no parametric distribution could be identified to fit the data correctly. As viral load is de facto nonnegative, a logarithmic transformation for near-zero values was used to correct

the estimate for adverse boundary effects [2, 3].

Additional details regarding HIV serological/nucleic acid testing : As mentioned in the manuscript, performance, limit of detection (LoD) and system failure (Sf) parameters were chosen based on routinely used procedures in Canadian blood operators. Specifically, HIV 1/2 antibody detection was modelled according to characteristics of Abbott PRISM HIV O Plus (Abbott Laboratories, Abbott Park, IL) for use on PRISM nEXT. Confirmatory testing for HIV antibodies was simulated based on the features of Geenius HIV 1/2 assay (Bio-Rad Laboratories, Marnes-la-Coquette, France) used at CBS and the ones of Abbott PRISM HIV O Plus assay utilized at Héma-Québec. HIV nucleic acid was designed according to their performance in 6 unit mini-pools using Roche cobas®MPX for use on the cobas®6800/8800 Systems (Roche Molecular Systems, Branchburg NJ) employed at CBS and the one in 16 unit mini-pools using Procleix Ultrio Plus (Grifols Diagnostic Solutions, Emeryville, CA) used at Héma-Québec. NAT LoD and Sf were obtained from Cobas multiplex and Procleix Ultrio Plus package inserts [4, 5]. The highest error between the two tests was used. A conversion factor of 0.62 was used to convert the LoD from IU/mL to copies/mL.

Parameter values associated to NAT performance and PI log reduction used in the simulations are displayed in Data S1 Table 2.

Additional details regarding PI steps : PI for plasma-derived products involves several methods such as pasteurization [6]; solvent/detergent (S/D) treatment [7]; Dry heat treatment of lyophilized products [8]; and Vapor heat [9].

REFERENCES

1. Hologic, Inc. Aptima™ HIV-1 Quant Dx Assay-Package Insert [Internet]. Hologic, Inc.; 2017 [cited 2021 May 26]. Available from : www.hologic.com/sites/default/files/package-insert/AW-11853-00100301.pdf
2. Geenens G, Wang C. Local-Likelihood Transformation Kernel Density Estimation for Positive Random Variables. *Journal of Computational and Graphical Statistics*. 2018 ; 27 :822-835.
3. Charpentier A, Flachaire E. Log-transform kernel density estimation of income distribution. *L'Actualité économique*. 2015 ;91 :141-159.
4. Cobas MPX test package insert. [Accessed Dec 16, 2020].
URL : www.fda.gov/media/115031/download
5. Procleix Ultrio Plus assay package insert. [Accessed Dec 16, 2020].
URL : www.fda.gov/media/73892/download
6. World Health Organization. Annex 4. Guidelines on viral inactivation and removal procedures intended to assure the viral safety of human blood plasma products. WHO Technical Report, Series No. 924. Geneva, Switzerland ; 2004 :150-224. [Accessed Oct 8, 2020]. URL : <https://www.who.int/bloodproducts/publications/>
7. Solheim BG. Pathogen reduction of blood components. *Transfus Apher Sci*. 2008 Aug ;39 :75-82. Epub 2008 Jul 3. PMID : 18602343.
8. Roberts PL, Dunkerley C, McAuley A, et al. Effect of manufacturing process parameters on virus inactivation by dry heat treatment at 80 °C in factor VIII. *Vox Sanguinis*. 2007 ;92 :56-63.

9. Bebulin (Factor IX Complex). Westlake Village, CA : Baxter Healthcare Corporation ; 2011.
10. Germain M, Remis RS, Delage G. The risks and benefits of accepting men who have had sex with men as blood donors. *Transfusion*. 2003 ;43 :25-33

Appendix S1 Table 1: Additional proportions used in the simulation¹

		3-month scenario (95%CI)	No deferral scenario (95%CI)
Donor proportion by age group	18-24	0.0827 (0.0693 – 0.0964)	
	25-44	0.2319 (0.2160 – 0.2517)	
	45+	0.6854 (0.6579 – 0.7057)	
Proportion of male donors by age group	18-24	0.5825 (0.5421 – 0.6362)	0.5827 (0.5423 – 0.6364)
	25-44	0.6586 (0.6169 – 0.7003)	0.6588 (0.6171 – 0.7004)
	45+	0.7667 (0.7481 – 0.8004)	0.7669 (0.7483 – 0.8005)
Proportion of donations made by new donors	18-24 Men	0.0448 (0.0255 – 0.0804)	
	Women	0.0564 (0.0346 – 0.0809)	
	25-44 Men	0.0178 (0.0119 – 0.0228)	
	Women	0.0317 (0.0162 – 0.0434)	
	45+ Men	0.0044 (0.0018 – 0.0075)	
	Women	0.0123 (0.0066 – 0.0208)	

¹Beta distribution was used for simulated proportions

Appendix S1 Table 2 – Parameter values for NAT performance and PI log reduction

Parameter	Value	Reference
Viral load (copies/mL)		Based on HIV viral load of infected donation data from United States database
Mean (95% CI)	99,197.85 (70,219.5 – 128,176.5)	
Median [Q1, Q3]	14,684.5 [3045.75 – 58643.25]	
NAT Limit of Detection (IU/mL) (95% CI) ¹	21.2 (18.2 – 25.7)	Procleix Ultrio Plus Package Insert [4]
NAT System failure (95% CI) ¹	0.00% (0.00 – 1.22)	Cobas MPX Package Insert [3]
Test transcription error (95% CI) ²	1:10,000,000 (1-100,000,000 – 1:4,000,000)	Germain et al. [9]
Clinical test error (95% CI) ¹	1:100,000 (1:1,000,000 – 1:35,500)	Germain et al. [9]
PI log-reduction	9.6	CSL Behring Statement

¹Variability modelled by a beta distribution

Data S2

Sensitivity Analysis

In order to assess the relative importance of key parameters in the residual risk estimation, we conducted a one-parameter-at-a-time sensitivity analysis. We considered the parameters associated with MSM HIV prevalence, MSM HIV incident donation probability, NAT limit of detection (LoD) and system failure, which were in turns either double or quadrupled as compared to the baseline scenario. Analysis of the obtained results has led to the following conclusions.

1. Doubling HIV incident donation rate in MSM donors (based on the probability of getting an incident donation instead of the number of new HIV cases in the population in a given period of time) entails an increase in the number of window period (WP) donations and the number of pools with positive viral load.
2. Doubling HIV prevalence in MSM repeat donors yields an increase in the number of prevalent donations without significantly impacting the risk of getting an infectious pool.
3. Doubling HIV prevalence in MSM first-time donors has the effect of slightly increasing the number of prevalent donations and viral load concentrations.
4. Doubling or quadrupling the NAT system failure rate and LoD has virtually no effect on the number of infected donations nor on the residual risk.
5. A significant increase in viral load concentration could potentially be problematic in a scenario where the pathogen inactivation process would not be used, since pool viral concentrations could reach a high level, which could cause one or more copies of the virus to end up in a pool.
6. The impact of modifying the parameters related to noncompliance in men is negligible. The complete results are shown in Appendix S2 Table 1. In addition to the one-parameter-at-a-time sensitivity analysis scenarios discussed above, the results

below include cases involving variations over more than one parameter.

It is worth pointing out that, to the best of our knowledge, there is, so far in the literature related to HIV risk estimation involving MSM, no substantial history of including thorough assessments of model parameter effects on residual risk values. Most of the previous studies in this field have depicted various deferral models without implementing parameter variations [1, 2, 3]. As it is impossible to conduct a clinical trial to collect data to evaluate the HIV residual risk following a change in MSM donor deferral criteria, parameters in simulations are mostly estimated based on historical data and it is arguably important to evaluate the effect of parameter variations. Most likely, optimistic and first pessimistic scenarios were based on the scenarios considered in O'Brien et al. 2019 [4]. For the second pessimistic the goal was to model the highest possible donor proportion using MSM general population proportion.

REFERENCES

1. Germain M, Remis RS, Delage G. The risks and benefits of accepting men who have had sex with men as blood donors. *Transfusion*. 2003 ;43 :25-33.
2. Anderson SA, Yang H, Gallagher LM, et al. Quantitative estimate of the risks and benefits of possible alternative blood donor deferral strategies for men who have had sex with men. *Transfusion*. 2009 ;49 :1102-1114.
3. Davison KL, Conti S, Brailsford SR. The risk of transfusion-transmitted HIV from blood donations of men who have sex with men, 12 months after last sex with a man : 2005-2007 estimates from England and Wales. *Vox Sang*. 2013 ;105 :85-88.
4. O'Brien SF, Grégoire Y, Pillonel J, et al. HIV residual risk in Canada under a three-month deferral for men who have sex with men. *Vox Sang*. 2020 ;115 :133-

139.

Appendix S2 table 1. Sensitivity analysis – Results based on parameter changes

MOST LIKELY (3-month)	HIV prev. in non-compliant MSM first time donor doubled ¹	HIV prev. in MSM first time donor doubled ¹	HIV prev. in non-compliant MSM repeat donor doubled ¹	HIV prev. in MSM repeat donor doubled ¹	HIV incidence in non-compliant MSM repeat donor doubled ¹	HIV incidence in MSM repeat donor doubled ¹	NAT LoD		NAT system failure probability	
							(x2) ¹	(x4) ¹	(x2) ¹	(x4) ¹
HIV positive donations per 1,000,000 donations	2.79	2.82	2.91	2.87	3.52	2.79	2.72	2.79	2.79	2.79
Number of pools with a viral load (in 300,000 pools)	1323	1302	1314	1236	1260	1323	1696	1323	1323	1323
Probability of getting an infected pool	0.00441	0.00434	0.00438	0.00412	0.00420	0.00441	0.00565	0.00441	0.00441	0.00441
Mean number of copies per pool after NAT and PI ($\times 10^{-3}$ copies)	9.7606	10.138	26.636	181.49	9.3127	9.7606	9.2143	9.7606	9.7606	9.7606
Median number of copies per pool after NAT and PI ($\times 10^{-3}$ copies)	1.7536	1.8600	1.7450	1.5800	1.6350	1.7536	1.7350	1.7536	1.7536	1.7536
Maximum number of copies per pool after NAT and PI ($\times 10^{-6}$ copies)	1.7672	1.6600	147.39	213.23	2.0300	1.7672	1.7000	1.7672	1.7672	1.7672

¹ Compared to the most likely scenario of the 3-month deferral model.

BIBLIOGRAPHIE

1. Public Health Agency of Canada. HIV in Canada. Ottawa : Public Health Agency of Canada ; 2018. http://publications.gc.ca/collections/collection_2018/aspc-phac/HP40-216-2018-eng.pdf. Accessed August 7, 2019.
2. Weusten J, Vermeulen M, van Drimmelen H, Lelie N. Refinement of a viral transmission risk model for blood donations in seroconversion window phase screened by nucleic acid testing in different pool sizes and repeat test algorithms : VIRAL TRANSMISSION RISK IN WINDOW PHASE. *Transfusion*. 2011 ;51(1) :203-215. doi :10.1111/j.1537-2995.2010.02804.x
3. Weusten J, Van Drimmelen HAJ, Lelie PN. Mathematic modeling of the risk of HBV, HCV, and HIV transmission by window-phase donations not detected by NAT. *Transfusion*. 2002 ;42(5) :537-548. doi :10.1046/j.1537-2995.2002.00099.x
4. Cappy P, Barlet V, Lucas Q, et al. Transfusion of HIV-infected blood products despite highly sensitive nucleic acid testing : HIV-INFECTED BLOOD DESPITE ID-NAT. *Transfusion*. 2019 ;59(6) :2046-2053. doi :10.1111/trf.15203
5. Germain M, Remis RS, Delage G. The risks and benefits of accepting men who have had sex with men as blood donors. *Transfusion*. 2003 ;43(1) :25-33. doi :10.1046/j.1537-2995.2003.00281.x
6. Anderson SA, Yang H, Gallagher LM, et al. Quantitative estimate of the risks

- and benefits of possible alternative blood donor deferral strategies for men who have had sex with men. *Transfusion*. 2009 ;49(6) :1102-1114. doi :10.1111/j.1537-2995.2009.02124.x
7. Davison KL, Brant LJ, Presanis AM, Soldan K. A re-evaluation of the risk of transfusion-transmitted HIV prevented by the exclusion of men who have sex with men from blood donation in England and Wales, 2005-2007 : HIV risk of blood donations from MSM. *Vox Sanguinis*. 2011 ;101(4) :291-302. doi :10.1111/j.1423-0410.2011.01491.x
 8. Schlenke P. Pathogen Inactivation Technologies for Cellular Blood Components : an Update. *Transfus Med Hemother*. 2014 ;41(4) :309-325. doi :10.1159/000365646
 9. Irsch J, Lin L. Pathogen Inactivation of Platelet and Plasma Blood Components for Transfusion Using the INTERCEPT Blood System™. *Transfus Med Hemother*. 2011 ; 38(1) :19-31. doi :10.1159/000323937
 10. Comité de pilotage de l'enquête Complidion et de la surveillance épidémiologique des donneurs de sang. Bilan sur le respect des critères de sélection des donneurs de sang deux ans après leur modification. Santé publique France. Saint Maurice : Santé publique France, 2018. 10 p. Disponible à partir de l'URL : www.santepubliquefrance.fr
 11. O'Brien SF, Osmond L, Fan W, Yi Q-L, Goldman M. Compliance with time-based deferrals for men who have sex with men : 1-YEAR DEFERRAL FOR MSM. *Transfusion*. 2019 ;59(3) :916-920. doi :10.1111/trf.15098
 12. O'Brien SF, Grégoire Y, Pillonel J, et al. HIV residual risk in Canada under a three-month deferral for men who have sex with men. *Vox Sang*. 2020 ;115(2) :133-139. doi :10.1111/vox.12867
 13. Notice d'accompagnement du test Cobas MPX.
URL : <https://www.fda.gov/media/115031/download>

14. Charpentier A, Flachaire E. Log-transform kernel density estimation of income distribution. Published online 2014 :22.
15. Geenens G, Wang C. Local-Likelihood Transformation Kernel Density Estimation for Positive Random Variables. *Journal of Computational and Graphical Statistics*. 2018 ; 27(4) :822-835. doi : 10.1080/10618600.2018.1424636
16. MacDonald NE, O'Brien SF, Delage G. Transfusion and risk of infection in Canada : Update 2012. *Paediatr Child Health*. 2012 ;17(10) :e102-e106.
17. Wilson K, Atkinson K, Keelan J. Three decades of MSM donor deferral policies. What have we learned ? *International Journal of Infectious Diseases*. 2014 ;18 :1-3. doi :10.1016/j.ijid.2013.09.016
18. Davison KL, Conti S, Brailsford SR. The risk of transfusion-transmitted HIV from blood donations of men who have sex with men, 12 months after last sex with a man : 2005-2007 estimates from England and Wales. *Vox Sang*. 2013 ;105(1) :85-88. doi :10.1111/vox.12024
19. Germain M. The risk of allowing blood donation from men having sex with men after a temporary deferral : predictions versus reality. *Transfusion*. 2016 ;56(6pt2) : 1603-1607. doi :10.1111/trf.13541
20. Goldman M, W-Y Shih A, O'Brien SF, Devine D. Donor deferral policies for men who have sex with men : past, present and future. *Vox Sang*. 2018 ;113(2) :95-103. doi :10.1111/vox.12623
21. Davison KL, Gregoire Y, Germain M, et al. Changing the deferral for men who have sex with men—an improved model to estimate HIV residual risk. *Vox Sang*. 2019 ; 114(7) : 666-674. doi :10.1111/vox.12826
22. Germain M, Robillard P, Delage G, Goldman M. Allowing blood donation from men who had sex with men more than 5 years ago : a model to evaluate the impact

- on transfusion safety in Canada. *Vox Sang.* 2014;106(4) :372-375. doi :10.1111/vox.12109
23. Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ. Probabilistic Networks and Expert Systems. Springer ; 1999.
 24. Camirand Lemyre F, Notes de cours : STT760 - Math pour l'intelligence artificielle. 2020.
 25. Sheather SJ, Jones MC. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society : Series B (Methodological)*. 1991;53(3) :683-690. doi :10.1111/j.2517-6161.1991.tb01857.x
 26. Ford C. Understanding Q-Q Plots. | University of Virginia Library Research Data Services + Sciences. Published August 26, 2015. Accessed August 5, 2020. [https ://data.lib- ary.virginia.edu/understanding-q-q-plots/](https://data.library.virginia.edu/understanding-q-q-plots/)
 27. Saporta G. Probabilités, analyse des données et statistique. Editions Technip ; 2008.
 28. Razali NM, Wah YB. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. :14.
 29. Shapiro SS, Wilk MB. An Analysis of Variance Test for Normality (Complete Samples). Published online 2020 :22.
 30. Les mesures de sécurité en place. Accessed September 29, 2020. [https ://www.hema- quebec.qc.ca/sang/savoir-plus/mesures-securite/mesures-securite-en-place.fr.html](https://www.hema-quebec.qc.ca/sang/savoir-plus/mesures-securite/mesures-securite-en-place.fr.html)
 31. Document d'information - Analyse des dons de sang : dépistage du VIH | Canadian Blood Services. Accessed September 29, 2020. www.blood.ca/fr/notre-sujet/ressources-pour-les-medias/vih/document-dinformation-analyse-dons-sang-depistage-du-vih
 32. Les technologies de dépistage du VIH | CATIE - La source canadienne de renseignements sur le VIH et l'hépatite C. Accessed June 6, 2019. www.catie.ca/fr/feuillets-info/depistage/les-technologies-depistage-vih

33. Rich AJ, Lachowsky NJ, Sereda P, et al. Estimating the Size of the MSM Population in Metro Vancouver, Canada, Using Multiple Methods and Diverse Data Sources. *J Urban Health*. 2018 ;95 :188-195.
34. Challacombe L. The epidemiology of HIV in gay men and other men who have sex with men. *Catie Fact Sheet*. 2015. [Accessed 2020 Nov 29].
URL : <https://www.catie.ca/ga-pdf.php?file=sites/default/files/epimsm%20EN%202015%2002%2025b.pdf>
35. Hotchko M, Robert P. Recent market status and trends of fractionated plasma products. *Ann Blood*. 2018 ;3 :19-19.
36. Koller D, Friedman N. Probabilistic Graphical Models : Principles and Techniques. MIT Press ; 2009.
37. Document d'information - Les dons de plasma au Canada. Consulté le 7 juillet 2021. <https://www.canada.ca/fr/sante-canada/services/medicaments-produits-sante/participation-public-consultations/biologiques-radiopharmaceutiques-therapies-genetiques/document-information-dons-plasma-canada.html>
38. Homme ayant des relations sexuelles avec d'autres hommes (HARSAH). Consulté le 7 juillet 2021. <https://sang.ca/fr/sang/puis-je-donner/hommes-ayant-eu-un-rapport-sexuel-avec-un-homme-harsah>
39. Puis-je donner ? Consulté le 7 juillet 2021. <https://www.hema-quebec.qc.ca/sang/donneur-sang/puis-je-donner/index.fr.html>
40. Wilson K, Atkinson K, Keelan J. Three decades of MSM donor deferral policies. What have we learned ? *International Journal of Infectious Diseases*. 2014 ;18 :1-3. doi :10.1016/j.ijid.2013.09.016
41. Goldman M, Shih AW-Y, O'Brien SF, Devine D. Donor deferral policies for men who have sex with men : past, present and future. *Vox Sanguinis*. 2018 ;113(2) :95-103. doi :10.1111/vox.12623

42. Pillonel J, Pelat C, Tiberghien P, et al. The evolving blood donor deferral policy for men who have sex with men : impact on the risk of HIV transmission by transfusion in France. *Transfusion*. 2020 ;60(3) :525-534. doi :<https://doi.org/10.1111/trf.15677>
43. Revised Recommendations for Reducing the Risk of Human Immunodeficiency Virus Transmission by Blood and Blood Products; Guidance for Industry. :21. <https://www.fda.gov/media/92490/download>.
44. Moiz B, Moatter T, Shaikh U, et al. Estimating window period blood donations for human immunodeficiency virus Type 1, hepatitis C virus, and hepatitis B virus by nucleic acid amplification testing in Southern Pakistan. *Transfusion*. 2014 ;54(6) : 1652-1659. doi :10.1111/trf.12521
45. Chapelin K, Grenier E. Comment construire un diagramme de Henry avec Excel et comment l'interpréter. *Revue Romuald*. 2006. <http://www.modulad.fr/numero-35/Excelense-grenier-35/DiagrammedeHenry.pdf>
46. Silverman BW. *Density Estimation for Statistics and Data Analysis*. London : Chapman & Hall, 1986.
47. Hall P. Cross-Validation in Density Estimation. *Biometrika*. 1982 ;69(2) :383-390. doi :10.2307/2335412
48. Evans, M., Rosenthal, J. S. (2010). *Probability and statistics : the science of uncertainty*, second edition. New York : W.H. Freeman and Co. Ce livre est disponible en distribution gratuite non destinée à la vente à l'adresse <http://www.utstat.toronto.edu/mikevans/jeffrosenthal/>
49. Scott DW, Terrell GR. Biased and Unbiased Cross-Validation in Density Estimation. *Journal of the American Statistical Association*. Published online March 12, 2012. <https://www.tandfonline.com/doi/abs/10.1080/01621459.1987.10478550>
50. Devroye L, Györfi L. *Nonparametric Density Estimation*. Wiley Series in Probability and Mathematical Statistics. New York : John Wiley & Sons. 1985.

<https://www.szit.bme.hu/~gyorfi/L1bookBW.pdf>.